DOCUMENT RESUME

ED 192 166 CE 027 161

AUTHOR Williamson, Stephen A.: Schaalman, Mary Lou
TITLE The Assessment of Occupational Competence. 2. Assessment Centers: Theory, Practice, and Implications for Education.
INSTITUTION McBer and Co., Boston Mass.
SPONS AGENCY National Inst. of Education (DHEW), Washington, D.C.
PUB DATE Feb 80
CONTRACT 400-78-0028
NOTE 301p.: For related documents see CE 027 159-165.

EDRS PRICE MF01/PC13 Plus Postage.
DESCRIPTORS Adults: *Competence: Employment Qualifications: Job Performance: *Job Skills: *Minimum Competency Testing: Occupational Tests: *Personnel Evaluation: *Personnel Selection: Predictive Measurement: Vocational Aptitude
IDENTIFIERS *Occupational Competence Assessment: *Testing Centers

ABSTRACT

Focusing on assessment centers, this chapter is one of seven in a report that examines the assessment of occupational competence. Assessment centers are presented first in a historical context (their origins in personality theory), through which their evolution is traced and examined. What follows is a description of current assessment center practice, involving a broad sample of presently active assessment centers. The third major portion discusses the process and the results of the assessment center to illuminate the following key issues for the researcher and practitioner: Is the assessment center useful as a method of ascertaining the presence or absence of specified competencies? Does it measure performance potential or current ability? To what use do organizations put the data that they gather on individuals? What are the consequences of implementing assessment centers for the persons who are assessed and for the implementing organizations? The chapter concludes with an examination of the functions assessment centers serve for the organizations that use them. (Other sections of the report are available separately--see note. The first is an overview: the last is a synthesis of issues.) (LRA)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED192166

CE 027 161

THE ASSESSMENT OF OCCUPATIONAL COMPETENCE

2. ASSESSMENT CENTERS: THEORY, PRACTICE, AND IMPLICATIONS FOR EDUCATION

Stephen A. Williamson
Mary Lou Schaalman

Prepared for
The National Institute of Education
Contract NIE 400-78-0028

February, 1980

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

McBER 137 NEWBURY STREET · BOSTON · MASSACHUSETTS 02116

DEC 4 1980

TABLE OF CONTENTS

I. Introduction 2.1

Report Purpose and Outline 2.1
What Is an Assessment Center? 2.3

II. Historical Foundations 2.14

Origins of the Assessment Center Concept 2.14
The OSS Assessment Center Program (1943-1945) 2.17
Post-War Applications and Problems 2.33
The Management Progress Study 2.37
The Coming of Age of the Assessment Center Concept (1960-1979) 2.51
The Lessons of History 2.63

III. Present State of the Art in Practice 2.71

Nature of the Sample 2.71
Reasons for Setting Up an Assessment Center 2.75
Resources Required for Assessment Center Design 2.79
Purposes of the Assessment Center 2.80
Behavioral Dimensions Utilized 2.85
Design of Performance Situations 2.92
The Composition of Assessor Teams 2.95
Assessor Training 2.98
Methods for Selecting Assessees 2.103
Assessor Observations, Ratings, and Competency Judgments 2.106
The Feedback of Assessment Center Results 2.112
Further Consequences of Assessment Centers for Participants 2.117
Consequences of Assessment Centers for the Organization 2.124
Conclusions and Implications for Educational Applications 2.133

IV. Research Foundations 2.182

Are Assessment Centers Valid? 2.183
Are Assessment Centers Reliable? 2.216
What Do Assessment Centers Really Measure? 2.221
Are Assessment Centers Equitable? 2.227
Are Assessment Centers Cost-Effective? 2.229
Conclusions 2.232

V. Implications for the Educational Community 2.236

The Assessment Center Concept: Measuring Occupational Competence? 2.236
Assessment Centers in Education: Is There a Need? 2.243
The State of Practice: A Problem for Education? 2.252
The Research Base: Further Research Needed? 2.258

Appendix: Standardized Performance Situations

References

I. INTRODUCTION

Report Purpose and Outline

No discussion of the current state of the art in competence assessment would be complete without an examination of the phenomenon known as the "assessment center." In both private and public sector organizations, the decade of the 1970's has been witness to a remarkable growth in the use of assessment centers for the appraisal, selection, and development of human resources. Today, the assessment center technique is being actively considered or implemented in a variety of educational settings as a means for the evaluation of student and/or faculty performance. Thus, the educational community is in need of accurate information regarding the current state of the art of the assessment center concept. Furthermore, the interested educator needs a framework with which to assess the feasibility and desirability of applying the assessment center concept to the attainment of educational goals. The objective of this document is to meet these needs in a balanced and constructive fashion.

This report is the culmination of a year-long study of the assessment center phenomenon with respect to historical antecedents, theoretical foundations, current practice, and research data. We believe this picture of the current state of the art is of value per se. However, we have also attempted to pinpoint some fundamental questions and concerns for educators, researchers, and practitioners as the result of this review. This discussion of the relative strengths and weaknesses of the current state of the art is intended specifically to help the interested educator as well as to contribute more generally to the advancement of the field of competence assessment.

To meet our objectives, we have presented this report in five sections. The remainder of this introductory section is devoted to a definition and overview of the assessment center concept and its component techniques. This discussion should enable the reader to discern the unique approach to competence assessment represented by the assessment center concept.

The second section is a delineation of the history of the assessment center concept from its military origins during the World Wars to the factors underlying the rapid expansion of applications during the early '70's. The major theoretical foundations of the assessment center concept are isolated for discussion, and several recurrent themes are critically examined as the lessons of history for future applications.

The third section is a compilation of our study of current applications of assessment centers for various purposes in business, governmental, and educational settings. This review of the present state of the art in practice provides the reader with a practical guide to the issues and problems which arise in the conduct of an assessment center from its original conception of purpose and design to its operation and consequences.

The fourth section is a presentation and critical analysis of the major research studies of assessment centers. This research base is examined with respect to what human characteristics are measured in assessment centers; how these characteristics are measured; and, of course, the validity of these measurements in predicting important outcomes. Several fundamental problems and issues are noted regarding this literature.

Finally, the fifth section is addressed to the needs of the education community identified previously. The major conclusions reached in this study are reviewed and summarized. Key strengths and weaknesses in the current state of the art are assessed with respect to future educational applications; and several recommendations are identified for future research and practice.

What Is an Assessment Center?

The seemingly simple task of defining the assessment center concept is complicated by a rather more complex reality. The assessment center concept has appeared in the literature since the early 1940's, and quite a few definitions have been offered since then. Many practitioners have told us that the label is often misapplied to certain activities which are no more than the administration of a battery of tests to groups of individuals. (For example, one so-called "assessment center" we examined was, in fact, simply a battery of psycho-motor tests administered to individuals for entry into lower-level jobs.

Few practitioners would label this activity as within the domain of assessment centers.)

Writers in this field have used a variety of descriptions to define the assessment center concept. An assessment center has been referred to as a place, a facility, a method, a program, or a process. At the 1979 International Congress on the Assessment Center Method, one speaker suggested that an assessment center is now best defined as a technology. Indeed, assessment centers today appear to be all of these things in one way or another.

Douglas Bray, who is often referred to as the father of the modern assessment center movement, has offered perhaps the best definition of the concept. He suggests that the term, asssessment center, is:

> ...somewhat of a misnomer, since it implies that there must be a building or some other semi-permanent physical location for the activity. Although this is often the case, there is nothing mandatory about it. What is really involved in assessment is the application of various methods of observing and evaluating behavior in a variety of situations. The methodology is what is important, not the physical setting. The purpose of an assessment center is to provide an objective off-the-job evaluation of the developed abilities, potential, strengths and weaknesses, and motivation. The assessment center leads to these evaluations through the observation of behavior in a variety of standardized performance situations, the rating of that behavior on a number of pre-determined dimensions, drawing of conclusions concerning potential for certain levels and types of work, and the diagnosis of developmental needs (Bray, 1976).

Fortunately, a more definitive description of the assessment center concept has been developed as a result of a collaborative effort by many of the top practitioners in the field. This descriptive set of standards is at the heart of a document developed by a task force led by Joseph L. Moses of AT&T. This document, entitled "Standards and Ethical Considerations for Assessment Center Operations," was constructed in 1975, published in 1977 (Moses & Byham), and revised in 1979. We quote directly from it:

> To be considered as an assessment center, the following minimal requirements must be met:
>
> - Multiple assessment techniques must be used. At least one of these techniques must be a simulation. A simulation is an exercise or technique designed to elicit behaviors related to dimensions of performance on the job by requiring the participant to respond behaviorally to situational stimuli. The stimuli present in a situation parallel or resemble stimuli in the work situation. Examples of simulations include group exercises, in-basket exercises, and fact finding exercises.
>
> - Multiple assessors must be used. These assessors must receive training prior to participating in a center.
>
> - Judgements resulting in an outcome (i.e., recommendation for promotion, specific training or development) must be based on pooling information from assessors and techniques.
>
> - An overall evaluation of behavior must be made by the assessors at a separate time from observation of behavior.
>
> - Simulation exercises are used. These exercises are developed to tap a variety of pre-determined behaviors and have been pre-tested prior to use to ensure that the techniques provide reliable,

objective and relevant behavioral information for the organization in question.

- The dimensions, attributes, characteristics or qualities evaluated by the assessment center are determined by analysis of relevant job behaviors.
- The techniques used in the assessment center are designed to provide information which is used in evaluating the dimensions, attributes or qualities previously determined.

In summary, an assessment center consists of a standardized evaluation of behavior based on multiple judgements. Multiple trained observers and techniques are used. Judgements about behaviors are made, in part, from specially developed assessment simulations. These judgements are then pooled by the assessors at an evaluation meeting during which all relevant assessment data are reported and discussed, and the assessors agree on the evaluation of the dimensions and any overall evaluation that is made.

The following kinds of activities do not constitute an assessment center:

- Panel interviews or a series of sequential interviews as the sole technique;
- Reliance on a specific technique (regardless of whether a simulation or not) as the sole basis for evaluation;
- Using only a test battery composed of a number of pencil and paper measures, regardless of whether the judgements are made by a statistical or judgemental pooling of scores;
- Single assessor assessment (often referred to as individual assessment) - measurement by one individual using a variety of techniques such as paper and pencil tests, interviews, personality measures or simulations;
- The use of several simulations with more than one assessor where there is no pooling of data - i.e., each assessor prepares a report on performance in an exercise, and the individual reports (unintegrated) are used as the final product of the center;

- A physical location labelled as an 'assessment center' which does not conform to the requirements noted above.

These definitions provide a more-than-adequate focus for this report. We do not believe it is necessary to provide yet another definition of the assessment center concept. In fact, is seems quite appropriate to describe the assessment center concept as a technology. Webster's Dictionary defines technology as a "technical method for achieving a practical purpose." Assessment centers are used for achieving such practical purposes as selection, promotion, training, development, and career planning. The essence of the technical method (and its most distinguishing feature) is the foundation of the assessment in the measurement of multiple characteristics of the individual via multiple measurement tools involving at least some observation of the individual's behavior by multiple observers.

An assessment center requires five technical components for its operation. The first component is a set of behavioral dimensions. Target jobs for which individuals are to be assessed are described in terms of a limited set of descriptors reflecting the kinds of behaviors required by the job. These dimensions are usually, though not always, constructed from some type of job analysis. These dimensions are critical in that they represent the characteristics which must be observed and rated by observers.

The second component of the method is a set of standardized performance situations in which all individuals who are being assessed must be engaged. These situations can, and often do, include paper-and-pencil measures of psychological traits, simulations of target job activities, social interaction situations, and various types of interviews. The situations are designed to elicit the behavior represented by the behavioral dimensions. (Examples may be found in the appendix.)

The third component of the method is assessee behavior. A limited group (usually 6-12) of participants (hereinafter called assessees) are brought together at an off-the-job site, in groups, to respond to the performance situations. Assessees are free to behave as they see necessary to perform in the situations. Thus, variations in assessee behavior can be examined by observers within the context of relatively standard stimuli reflective of the job demands.

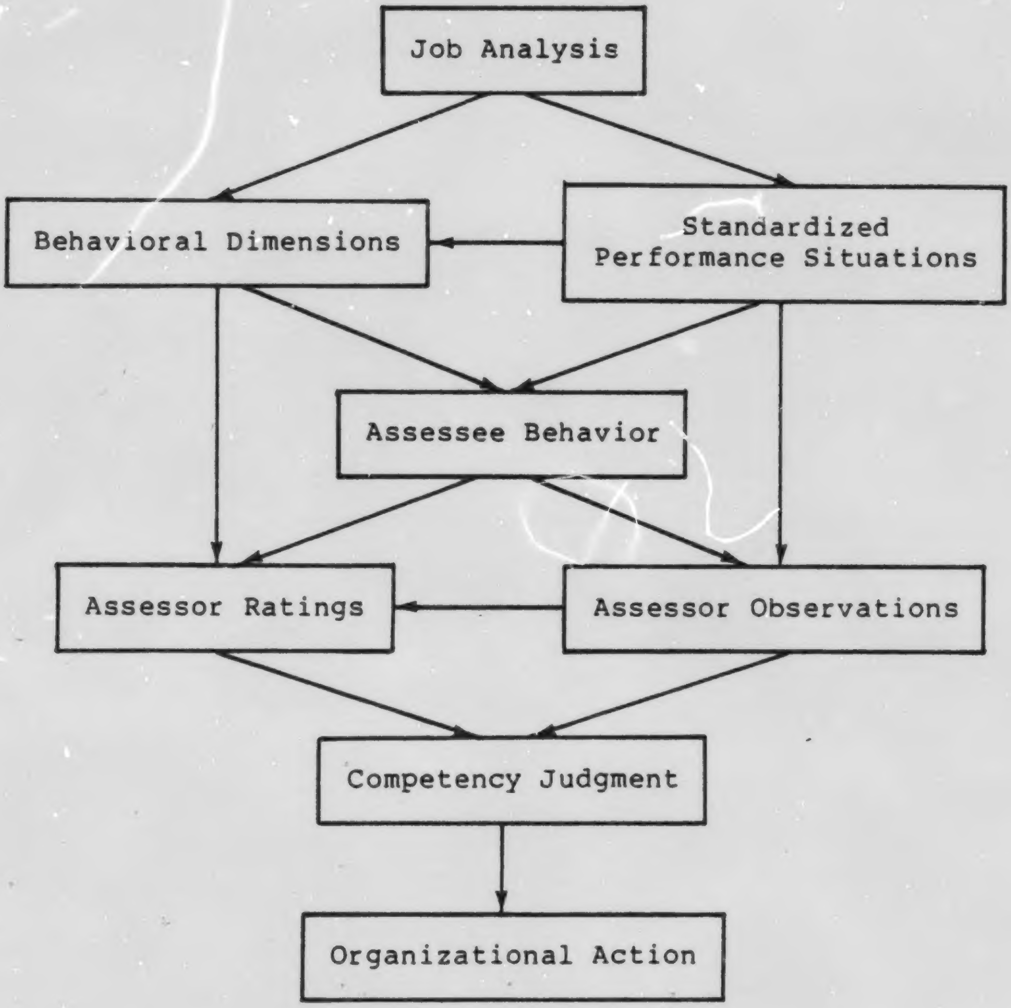
The fourth component of the method is a series of assessor observations. A limited group (usually 3-6) of observers, typically drawn from one to two organization levels above the target job, are trained to observe and record assessee behaviors in the performance situations. Within each performance situation, the assessors observing a particular assessee are asked to relate the assessee behaviors to a specific behavioral dimension and to rate (on some type of numerical scale) the quality of that behavior with respect to the dimension. (This rating procedure sometimes does not occur until the conclusion of the assessment center.)

The final component of the method is competency judgments. At the conclusion of the assessment center, the assessor team (usually) combines their ratings of specific assessees in different performance situations to form pooled ratings of each assessee along each behavioral dimension. In addition, the assessor team forms a pooled judgment (usually called an overall rating, or OAR) regarding each assessee's ability to perform the desired job. This judgment also is typically placed on a numerical scale. This summary competency judgment is then utilized with respect to the purpose of the center (e.g., recommendation for selection or promotion, or as an index of personal strengths and weaknesses for developmental purposes). Depending upon the purposes of the center, these judgments, and associated ratings, may or may not be shared with the assessee or other managers in the organization.

Figure 1 presents a schematic overview of the assessment center concept as just discussed. While this diagram simplistically represents the core components of the assessment center technology, it is important to emphasize the complex nature of this type of competence assessment process. Problems in concept or design in any of the components can lead to serious errors in the reliability or validity of the ultimate competency judgments. Far too many practitioners have described to us the difficulties that result from the simple lack of recognition of this complexity. It is not a technology for use by the beginner.

FIGURE 1

Schematic of the Technical Components
of the Assessment Center Concept


Job Analysis
Behavioral Dimensions
Standardized Performance Situations
Assessee Behavior
Assessor Ratings
Assessor Observations
Competency Judgment
Organizational Action

Despite this complexity (or perhaps because of it), the assessment center concept is quite consistent with recent trends in pyschological theory, and is quite compatible with most new practices in assessment and measurement technology. Thus, we believe that the concept of the assessment center does represent a significant advance in the application of the theory and practice of competence assessment along, at least, the following lines:

- It describes jobs with respect to required behaviors for effective performance, usually based on some a priori research on actual job performance of incumbents;
- It utilizes multitrait, multimethod assessment designs for reaching judgments about specific individual skills, knowledge, or characteristics;
- It utilizes believable simulations of job-related situations as a means of eliciting behavior for assessment;
- It bases judgments of individuals more on directly observable behavior than on self-report or paper-and-pencil measures;
- It requires the systematic pooling of multiple observers to reach competency judgments;
- It includes in many cases, a research design for an assessment of the quality of measurement; and
- It enables judgments of individual competence within work settings to be formulated in a more systematic fashion.

These are extremely important advances in the state of the art of competence assessment; therefore, the educational community would do well to take heed of them for educational goals. Yet, as we shall discuss throughout the remainder of this report, the interested educator must be aware of some current difficulties in applications of the assessment center concept. At this point we would identify the following issues.

- Much of the development of the technology has been driven by the demands of practice. While this is not inherently a problem, the field is in need of greater theoretical development. For example, there is a certain lack of clarity in the concept of behavioral dimensions. For many practitioners, this concept is a label for some aggregation of behaviors demanded by the job; yet, inferences regarding the underlying competencies of the individual are usually drawn. The relationship of observable behavior and individual competence is not clearly articulated. This theoretical problem has serious implications for the manner in which assessors translate observations of behavior into judgments of individual capacity to perform a job.

- As mentioned previously, the component technology is complex. Wide variations exist in practice, often to the detriment of particular centers. Many practitioners lament the fact that too many users misapply the concept, or ignore critical technical needs (e.g., the lack of systematic job analysis, quality observer training, or adequate design of performance situations).

- Certain components of the technology can, and should, be improved through the application of research and theory from other settings. For example, much of the research available on systematic observation techniques has not been utilized in the design of observation formats or in assessor training. Also, group dynamics theory might well provide insight into the sources of error in assessor discussions and ratings.

- Research results obtained on the technology, while generally positive, are still of mixed effect and varied quality. Significant methodological problems are apparent in this research. These problems need to be more effectively understood if the field is to advance.

- The application of the technology for such purposes as selection, training, or career development strikes at the heart of the relationshilp between the individual and the organization. Since the technology often presents a significant departure from past practice, serious organizational problems can arise if the change is not effectively managed. Such repercussions are even more likely, we suspect, in educational settings.

The weighing of the relative advantages and disadvantages of the assessment center concept forms the foundation for the remaining sections of this report. In summary, it is fair to say that the assessment center concept, at this point in time, is reasonably well established, represents a significant advance in competence assessment practice, and should be actively examined by the educational community. Nevertheless, our analysis of the state of the art leads to one inescapable caution to the reader: despite the advances represented conceptually in assessment center technology, the state of the art has yet to reach its full promise. Much more research and development is needed in the next few years to move the technology closer to acceptable standards of validity and practical feasibility in educational settings. We are hopeful that this report will contribute to this advancement.

II. HISTORICAL FOUNDATIONS

Origins of the Assessment Center Concept

The origins of the assessment center concept can be traced to at least three sources. The first source was the work of Simoneit (1940, 1944), in Germany, between the World Wars. Though his work has yet to be published in English, Simoneit apparently developed a variety of multiple assessment techniques (including role plays and leaderless group discussions) for use by the German military for the selection of officers with respect to such traits as will power or rigidity.

A second origin can be found with the British adaptation of Simoneit's methods to the selection of Army officers in the early 1940's from the rank-and-file. Wartime conditions had necessitated the rapid expansion of the officer corps and had precluded the use of normal access channels (e.g., the schools of the upper class). This initial assessment center application was developed, in effect, to legitimate access to the officer corps in lieu of standard academic credentials.

Garforth (1945), a British colonel, summarized the operation of these War Office Selection Boards (WOSB) at the close of World War II. Each assessee was administered standard intelligence tests (unspecified) as well as five other paper-and-pencil psychological tests. Each candidate was also interviewed by a psychiatrist, a deputy of the WOSB (for "officer quality") and by another officer (for "technical qualifications"). However, the heart of the program was a series (exact number unspecified) of standardized performance situations. These situations included individual role plays (e.g., interviewing an actor playing a subordinate, or delivering a morale speech); "command situations," where the assessee was required to lead a group to task accomplishment; and leaderless group situations (e.g., discussions, or task situations such as improvising an escape). A final board conference was held during which all officers who had observed or interviewed the assessee would discuss the candidate's suitability for officer selection. Candidates were rated on an 8-point scale (from "rejected" to "outstanding"). After discussion, the President of the Board was charged with making the final decision for acceptance or rejection.

In this brief description, one can see the nascent outlines of the modern assessment center concept. Interestingly enough, Cronbach (1970) cites research by Vernon and Parry (1949) on these boards. While criterion data was often lacking, they did isolate one sample of 500 officers rated in combat by their subordinates. A moderate correlation (.35*) was found between these combat ratings and ratings by the WOSB. While reliability problems were uncovered in one study of two boards assessing the same subjects, reliability was substantially improved through systematic training of the boards in "similar standards and procedures" (Cronbach, 1970).

---

* A further discussion of the meaning of these statistics can be found in Section IV on research data.

A third origin of the assessment center concept can be traced to the work of Henry Murray and his colleagues at the Harvard Psychological Clinic in the 1930's. Murray's group was far less concerned with the practical problems of officer selection; rather, they had undertaken the ambitious enterprise of building a theory of personality from extensive assessment of some 51 college and noncollege students. The results of this effort were published in a now-classic text entitled Explorations in Personality (Murray, 1938).

Each subject was called upon to complete a wide range of tests and procedures, including written autobiographies, interviews on a variety of topics (e.g., family, sexual development, childhood memories, etc.), questionnaires and attitude surveys, ability tests, hypnotic tests, free association techniques, projective tests (such as the Rohrschach and TAT), and various "experimental" situations, such as "staged" interactions or situations. The goal of Murray's project was no less than the best possible formulation of the complete personality of each subject.

From the viewpoint of this report, the critical contribution of Murray's text was the conceptual foundation that he attempted to construct to account for the data generated by the host of assessment techniques. Murray argued that the personality of the individual is best described as the result of complex combinations of observable situational demands and unobservable, underlying needs of the individual. These combinations produce the observable behavior from which we infer the personality of the individual. By systematically varying the nature of the situation, one could observe the resultant behavior and assess the nature of the underlying needs of the individual.

In this formulation (decidedly oversimplified here), Murray offered one theoretical rationale for the design of the assessment center concept. This rationale was that prediction of future behavior could best be made by relying heavily on psychodynamic inferences of unobservable constructs of the personality drawn from observable behavior in multiple defined situations. The onset of World War II enabled the practical experiences of the Germans and the British and the theoretical studies of Murray to be combined for a massive initial application of the assessment center concept in the United States.

The OSS Assessment Center Program (1943-1945)

During the summer of 1943, an official of the newly created U.S. Office of Strategic Services (OSS) observed the operation of one of Britain's War Office Selection Boards. Faced with the equally difficult, if not impossible, task of selecting individuals for secret, behind-the-lines operations, this official convinced the OSS hierarchy to construct a similar selection program in the United States. By November of 1943, John Gardner, Joseph Gengerelli, and Henry Murray had been enticed to solve this problem. Barely two weeks later, the first assessment center in the United States was in operation.

This prodigious effort remains a classic case study of the creative development of the assessment center concept. Through the efforts of many top psychologists and other military personnel, this application laid the basic foundation for the assessment center concept and its application in the years to come. Thus, it is important to understand the basic thrust of this work. Nevertheless, only a brief overview can be provided here. The interested reader should consult the published results of this work, entitled Assessment of Men (1948), for further details. We shall examine the work in terms of the five components of assessment center technology outlined previously.

Behavioral Dimensions

The designers of this center operated under considerable constraints, not the least of which was the pressure of time under emergency wartime conditions. They had very little data (apart from some interviews with combat commanders) regarding the nature, and great variety, of the jobs to which assessees might be assigned. There were many foreigners and first-generation Americans in the sample of assessees, and many assessees ended up in assignments for which they had not been assessed. In addition, OSS recruitment practices, carried out under a veil of secrecy, tended to be attractive "to the bored, to the pathologically adventuresome, to those neurotically attracted to danger, and to pyschopaths in general" (MacKinnon, 1974).

Since the job of undercover agent could not be empirically defined in terms of specific required behaviors, the design staff resorted to the use of Murray's theoretical principles of "organismic assessment" to obtain a "picture of the whole man." Initially, some twenty general dimensions of personality were constructed for measurement. After several trial assessments, this list was refined to seven general and three specific dimensions of qualification (see Table 1).

These ten dimensions of personality, some of which were apparent in Murray's previous work, provided the conceptual framework for the assessment process. Even today it is difficult to argue with this selection of dimensions. However, without an empirical job analysis, the design staff could not be sure that these dimensions were truly important to job performance, nor could they concretely and systematically define the specific behaviors an assessee needed to display in order to recieve a high or low rating on any one dimension. This lack of specificity in the dimensions allowed the assessors considerable latitude in the interpretation of specific behaviors displayed by an assessee in any one performance situation, and required many inferential judgments

TABLE 1

OSS Assessment Center Behavioral Dimensions

General

1. Motivation for Assignment: war morale, interest in proposed job.

2. Energy and Initiative: activity level, zest, effort, initiative.

3. Effective Intelligence: ability to select strategic goals and the most efficient means of attaining them; quick, practical thought--resourcefulness, originality, good judgment--in dealing with things, people or ideas.

4. Emotional Stability: ability to govern disturbing emotions, steadiness and endurance under pressure, snafu tolerance, freedom from neurotic tendencies.

5. Social Relations: ability to get along well with other people, goodwill, team play, tact, freedom from disturbing prejudices, freedom from annoying traits.

6. Leadership: social initiative, ability to evoke cooperation, organizing and administering ability, acceptance of responsibility.

7. Security: ability to keep secrets; caution, discretion, ability to bluff and mislead.

Specific

8. Physical Ability: agility, daring, ruggedness, stamina.

9. Observing and Reporting: ability to observe and to remember accurately significant facts and their relations, to evaluate information, to report succinctly.

10. Propaganda Skills: ability to apperceive the psychological vulnerabilities of the enemy; to devise subversive techniques of one sort or another; to speak, write, or draw persuasively.

in reaching an overall rating of each dimension at the conclusion of the center. The OSS staff commented about these problems:

> ...the staff members were not forced to limit themselves to the observation of behavior which could be immediately translated into ratings of variables (i.e., dimensions), but were free to keep their eyes open for any actions which might be indicative of significant individual traits. This freedom, however, had certain disadvantages: it was responsible for certain confusions and inconsistencies. On the one hand, it was desirable to rate each performance of a candidate in its own right, independent of his performance in other situations, and the assessors consciously attempted to do this. On the other hand, as they followed a candidate from situation to situation, familiarized themselves with the scores he had made on other tests, and observed him in informal contact, their growing understanding of him necessarily influenced their perception and interpretation of the more subtle behavioral and expressive signs, and to that extent also influenced their ratings. Hence the scores in the situations scheduled late in the program were influenced by the cumulative evidence, rather than soley by the performance in the single test. This was the logical outcome of the attempt to arrive at a maximally comprehensive and meaningful personality picture (OSS staff, 1948:205).

The problems confronted by the OSS staff in the design of their behavioral dimensions are still remarkably salient for the modern application of the assessment center concept. Their ten assessment variables were generally conceived to be unobservable aspects of the individual's underlying personality. Observed behavior in the performance situations required assessor interpretation, in light of all other behavior across the situations, before ratings of dimensions were made. This use of underlying constructs of personality, and the treatment of behaviors as indicators of the underlying constructs, has largely fallen into disfavor today in the assessment center field. The modern approach is to view the dimensions as aggregates of similar behaviors required by the job. Nevertheless, modern assessor teams still face the problem of assigning specific behaviors to the dimensions, and summarizing their ratings of dimensions across multiple performance situations. In short, some degree of inference (or what social scientists call clinical judgment) must be exercised by the assessor team, both individually and collectively; and therein lies a very great potential source of assessment error. It is because of this issue that top practitioners today devote considerable effort to the training of assessors in order to minimize the degree of inference required and in order to systematize the inferences that must be made. We shall return to this issue later in this report.

Standardized Performance Situations

To measure their behavioral dimensions, the OSS staff selected a variety of available methods for use, and constructed a number of others. Within the 3-1/2 day assessment process, assessees were subjected to some 35 different procedures. The interested reader should consult the OSS report to gain the full flavor of this remarkable battery of procedures.

The types of procedures used included:

(1) Standardized tests of intelligence, e.g., Otis Test of Mental Ability, Bennett Mechanical Comprehension, Signal Corps Aptitude Test.

(2) Projective tests, e.g., Thematic Apperception Test, Sentence Completion Test, and the Rorschach.

(3) Clinical personal history interview, and associated post-test interviews.

(4) Group situational tests, e.g., leaderless group situations (moving a group of men across a river; moving a group over two walls representing an escape path; and group discussions of selected topics); and assigned leadership problems in which one candidate was placed in charge of a group to solve some problem (killing an enemy mayor, blowing up a bridge, etc.).

(5) Individual situational tests, e.g., supervising two incorrigible workers on a construction project; or maintaining cover when placed in a high stress, interrogation situation; or interrogating an actor.

(6) Miscellaneous paper-and-pencil tests, e.g., health questionnaire, and survey of work preferences.

(7) Physical ability tests, e.g., an obstacle course, and athletic contests.

(8) Tests of observing and reporting, e.g., summarizing the characteristics of a man from belongings left in a room, or memory tests of maps, names, and faces.

(9) Propaganda skills tests, e.g., writing a propaganda piece for enemy workers; or identifying information needed to build a propaganda program in a particular setting; and, group debates.

(10) Improvised role plays, designed by the staff to create problematic situations for particular candidates about whom there was some question.

(11) Sociometric questionnaries and peer ratings.

(12) Casual observations at meals, rest periods.

Within this battery of procedures can be seen the antecedents of many of the procedures typically employed today.

The OSS staff did place greater emphasis on the use of paper-and-pencil measures and clinical interviews than would an assessment staff today. However, the observations of behavior from the many role plays and situational simulations were critical to their final judgments. Thus, the assessment center concept was firmly grounded in a multitrait, multimethod tradition based on behavioral observation.

Assessee Behavior

The 3-1/2 day assessment center was conducted at a country estate outside of Washington, D.C. Due to increased demands for assessment, the OSS staff designed one-day programs for two other sites, and several assessment centers were eventually set up overseas. Assessees attended the centers in groups of 18, and 3 such groups would be cycled through the center every 2 weeks. Assesses were told very little about the center prior to arrival, but were instructed that, due to the secrecy of the OSS operation, they would be required to assume and maintain a fictional identity for the duration of their stay at the center. Upon arrival, they were issued identical uniforms, given a brief introduction, and ushered quickly into the first test situation. Not surprisingly, each group of assessees typically developed fairly high morale in the center, since they were constantly engaged alone and as teams in a variety of challenging and novel situations. Descriptions of the center's operation are replete with comments concerning the degree to which the simulations were able to elicit behaviors corresponding to "real life." It was to these behaviors which the assessment staff attended for reaching judgments concerning the candidate's suitability for intelligence work overseas.

Assessor Observations

The staff of the center were divided into senior and junior members on the basis of age, psychological training, and practical experience. The junior staff were generally responsible for the administration of standardized tests, while senior staff conducted most of the interviews. Both groups participated in observing simulations, rating behaviors, organizing the results, and making final evaluations, though the senior staff were held more responsible for the final selection decision.

A team of two senior and one junior staff member was assigned to one subgroup of 5-7 assessees from each class of 18, and were responsible for all of the ratings of their assessees. Each test and/or performance situation was designed to provide ratings on 1 to 4 of the 10 major dimensions. All team members would observe the performance situations and independently rate their assessees on each designated dimension. Six-point rating scales were used (very inferior, inferior, low average, high average, superior, very superior). It became common practice for the staff to add pluses or minuses to these ratings, thereby creating, in effect, a 16-point scale. Since scale points were not really anchored to behavioral examples, the staff was instructed to bear in mind a near-normal distribution of ratings--though this distribution was "barely approximated in practice" (OSS staff, 1948:205).

Following the observations of performance situations, the staff team would meet to reach agreement on one set of dimensional ratings for each individual in that situation. Averaging was not allowed; consensus was required through discussion. The staff commented about this process:

> While it is true that occasionally the consensus achieved was a mere compromise, determined very frequently by factors in the raters' personalities, it is no less true that in many cases the discussions resulted in an interpretation of the candidate's behavior that was better documented, more detailed, and more consistent than the individual interpretations had been, and which was therefore genuinely accepted by all members of the team. In all cases the exchange of opinions served to inform each observer of the criteria of evaluation used by the other and thus contributed to a gradual formation of a frame of reference shared by the whole team. Thus the function of the rating conferences went far beyond that of merely obtaining one set of ratings. Frequently it led to the team's arriving, in a first approximation, at a conception of a candidate's personality, a conception which was to be checked and corrected through further observations (OSS staff, 1948:206).

By the end of the 3-1/2 days of assessment, each assessee had received anywhere from 2 to 10 ratings on each of the 10 dimensions. A huge scoreboard was placed in the staff conference room, on which was placed each candidate's ratings for each of the dimensions in each of the performance situations or tests. At this point, final judgments could be reached.

Competency Judgments

At mid-afternoon of the final day, all staff teams would meet in conference. The meeting would proceed with a discussion of each assessee. The senior staff interviewer would present a written personality sketch. A junior staff member (who came to be called the "situationalist") would present the behavior ratings from the performance situations. Other staff members contributed their impressions and "criticism, freely expressed, was abundant and widely varied" (OSS staff, 1948:212).

Once a personality sketch was approved, the staff turned to the task of reaching a final rating (once again on a 6-point scale) for each of the 10 dimensions. By consulting the scoreboard, the staff could examine the relative dispersion of ratings of each dimension across the nearly 35 different measures. In the case of little dispersion, the final rating could be derived by simple averaging; greater dispersion "called for a discussion of the reasons for this and for a decision as to which data should be given the greater weight" (OSS staff, 1948:214).

The final task of the staff conference was to reach a "fitness rating." Each staff member independently rated the fitness of the assessee for his or her assignment (six-point scale once again). These ratings were announced and a final overall fitness rating was obtained by averaging. "In similar fashion an overall rating was arrived at, representing an estimate of the total potentialities of the candidates for meeting the challenges of life--an exceedingly vague and difficult concept to define" (OSS staff, 1948:217). In the last few months, the job fitness ratings were transformed to a five-point scale of recommendation for selection (not recommended to highly recommended). A similar scale was used to predict fitness for work: (1) at rear or advanced bases, or behind enemy lines; (2) at different levels of responsibility; and (3) within 10 broad categories of jobs (e.g., administrator, operational agent, etc.). The personality sketch, ratings of the ten dimensions, and the overall fitness ratings were then provided to the OSS hierarchy by the next morning.

With minor differences, this procedure of pooling assessor ratings of dimensions across multiple measurement procedures, and thereby deriving some type of overall competency judgment, remains a significant aspect of modern practice. The OSS staff recognized the difficulties and potential sources of error in this process. These difficulties are still important considerations today.

For example, given the fact that each behavioral dimension is measured at least twice (if not more often) across the several performance situations, the assessor team (or the assessment center designer) must decide upon some rule for reaching one summary rating of each dimension for each assessee. The OSS staff noted that test procedures were often differentially weighted in the assessor team discussions to reach this overall rating of a dimension. This was done because of (a) the nature of the target job (when known); (b) some idea that certain tests measured certain dimensions better; or (c) the developing picture of the personality of the assessee (in the assessor team discussion) dictated a different configuration of test weights. In effect, the OSS staff began to develop the nascent outlines of a theory of measurement; unfortunately, this theory has not been systematically pursued in the years since.

An even greater concern is the manner in which the distinct ratings of each of the behavioral dimensions are used by the assessor team to generate the final overall rating, or competency judgment. The OSS staff relied on clinical judgment, available personality theory or practical hunches to reach their final "fitness ratings." They even freely admitted the possible bias due to "halo effects" (i.e., ratings on one dimension affecting ratings on other dimensions). What was lacking was a clear set of rules for summing the judgments across each of the dimensions. This is still a largely unsolved problem today.

Consequences of the OSS Assessment Center

The OSS effort also established a tradition in the assessment center field of conducting ongoing evaluations of the effectiveness of the assessment center process in predicting actual success on the job. Unfortunately, the OSS staff were once again highly constrained in the conduct of this research by the lack of rigorous ratings of assesses' actual job performance.

Despite these constraints, the staff managed to obtain four different sets of job performance ratings on a small sample of assesses who ultimately went overseas. These four ratings included: (1) overseas staff ratings based on interviews with the assessee's commanding officer and associates; (2) ratings of the assessee by the theater commander; (3) ratings of the assessee by staff at reassignment areas; and (4) ratings of the assessee by peers. All four ratings were subject to serious error; furthermore, they were not highly intercorrelated for the same assessee (a common result yet today among performance appraisals from sources with varying perspectives).

The correlations of the overall "fitness ratings" from the assessment center with these four criteria of actual job performance are shown in Table 2 (included here are the results of the one-day OSS programs created somewhat later).

Table 2*

Correlations of OSS Assessment Center Fitness Ratings and Job Performance Ratings

TYPE OF APPRAISAL	3-1/2 day Center		1-day Center	
	r	n	r	n
Overseas Staff	.37	88	.53	83
Peers	.19	93	.21	173
Theater Commander	.23	64	.15	158
Reassignment Area	.08	53	.30	178

*Adapted from OSS Staff (1948:423).

The correlations were generally low to moderate, with the exception of the overseas staff appraisal (considered the most valid by the OSS staff), for which the assessment center predicted moderately well. The correlations of assessment center ratings on each of the ten dimensions with ratings on these dimensions by field assessors were even lower (ranging from .06 to .32).

The OSS staff were generally disappointed with these results, even though the correlations with staff ratings of performance are reasonably close to the average result of more recent research. The results are even more remarkable in light of the serious design constraints in the research. A more recent re-analysis of this data by Wiggins (1973, reported in MacKinnon, 1974) suggests that the OSS assessment center increased the accuracy of selection by 14-18 percent. Once again, this initial application of the assessment center concept foreshadowed the difficulties in research that were to come, and which we shall discuss in the fourth section of this report.

Perhaps the more important consequence of the OSS effort was the final series of recommendations produced by the staff for future assessment center applications. In these recommendations can be found viable standards of practice for today. Several key issues were isolated by the OSS staff:

- The necessity for rigorous, pre-design job analyses, based on interviews and observations of job incumbents and observations of job incumbents and their associates, as a source for "personality determinants of success";
- The necessity for rigorous design of performance situations and measurement techniques for rating the dimensions;
- The necessity for designing the ratings of actual job performance as rigorously as and in concert with the assessment center ratings, prior to the operation of the center;
- The necessity to validate the assessment center based upon the prediction of actual job performance;
- The necessity to continuously formulate conceptual schemes of the individual personality; and
- The necessity to assess individuals for a long trial period without reporting ratings or decisions to the organization (in order to avoid biasing of the job performance ratings).

While enjoying mixed success, the OSS staff was optimistic about the future of the assessment center concept. They concluded:

> Despite the somewhat equivocal outcome of the whole enterprise, we believe that by trials and errors we have exposed and found remedies for most of the defects of the system, and have finally arrived at a plan as promising as any that is known to us for advancing the science of man, a plan which contains the necessary instruments for its own modification and perfection (OSS staff, 1948:493).

Post-War Applications and Problems

The decade following World War II failed to fulfill the promise of the assessment center concept as articulated by the OSS staff. There was no great rush to apply the techniques to assessment or selection practices in business settings, though Bass (1954) identified a growing application of the leaderless group discussion as an assessment technique, in government and industrial settings. However, a small number of psychologists attempted to use the assessment center concept in several large-scale research programs.

The general results of these research efforts were quite mixed. Vernon (1950) found that a 3-day assessment center, composed of a mixture of tests, interviews and performance situations, quite accurately predicted training success and initial job performance of British civil service candidates, as rated by their supervisors (average correlation = .41). Kelly and Fiske (1950) used a battery of tests, performance situations and peer ratings to predict the success of VA-trained clinical psychologists over 3-10 years. Correlations of assessment ratings with performance ratings were generally quite low, particularly for measures from the performance situations. Kelly concluded that the OSS style of assessment had doubtful validity, and cited the use of multiple assessors as problematic until sources of prediction error were more clearly understood. Holt and Luborsky (1958) reached similar conclusions at the Menninger Clinic when their battery of tests and interviews failed to predict meaningfully the success of psychiatrists. They were particularly concerned to find that while the assessors could accurately assess specific individual characteristics (i.e., behavioral dimensions), they were often unable to summarize these assessments into a meaningful whole for accurate prediction of future success--a problem noted previously in the OSS effort.

Finally, Holmen, et al. (1956) and MacKinnon conducted several large-scale studies of officer selection for branches of the U.S. military. Holmen found assessor ratings and performance situations predicted grades in Officer Candidate School fairly well (correlations = .25 - .55). MacKinnon discovered various problems with his criteria of officer effectiveness, and found that global predictions of success as an officer based on his assessment center data were generally quite low (correlations all .30 or below).

Thus, by 1955 there was a considerable dampening of enthusiasm for the assessment center concept within psychological circles. That year, Cronbach, in an address to the American Psychological Association, pronounced that "assessment in the

OSS style has now been proved a failure...because it involves hazardous inferences...using tests whose construct interpretations are dubious and personality theory which has more gaps than solid matter" (Cronbach, 1955).

Other writers were less convinced of the death of the concept. Bass (1954) examined the research available on the leaderless group discussion, an important technique in most assessment centers, and concluded that the technique was both reliable and valid. He further concluded that the reliability and validity of the technique was improved by minimizing assessor inferences of underlying personality constructs; by using the technique primarily to assess the specific ability to initiate structure in unstructured situations; and by considering such variables as group size and composition as having an effect on the results. In his approach, Bass signaled an important need to examine more closely the sources of error (and accuracy) in the assessment center process and its component measurement methods.

Taft's (1959) excellent review of the literature at the end of the decade brought into focus most of the critical problems emerging in the theory and research on the assessment center concept. Taft's conclusions are remarkably prescient of many problems apparent today, and can be summarized with respect to four of the components of assessment center technology:

- Behavioral dimensions: The use of underlying personality constructs is problematic. Such constructs need to be validated per se through more explicit (i.e., behavioral) definition in the assessment center and actual job situations. Since the impact of these constructs on actual performance may vary with the nature of the job situation, a more refined taxonomy of situations needs to be developed.

- Standardized performance situations: These procedures are too often designed or selected without any prior evidence that the procedure distinguishes effective performers presently on the job. Furthermore, the value of multiple data is problematic; it is not clear that additional data improves the ability of the assessor to predict future performances.
- Assessor observations: Highest validity is achieved by assessors familiar with the job and successful incumbents. Also, very little data exists to support the value of pooled ratings versus the judgment of the "best assessor."
- Competency judgments: Clinical summaries of multiple assessment procedures, or consensus-based group discussions by assessors may be no better, or even worse, than statistical summarization of the test results.

Taft also suggested that research studies attempting to validate the assessment center concept suffered from the unreliability of the criteria used for validation, and warned that arguments for the validity of performance situations based on their simulation of the job (i.e., content validity) was no substitute for validity based on forecasting future job performance (i.e., predictive validity).

Today, over 20 years later, most of these issues are still a subject of concern in the assessment center field. This is not to suggest that the last 20 years have brought no real progress. Rather, the debate engendered by the OSS-based assessment center concept was, to a certain extent, overshadowed by developments in practice outside the discipline of psychological research. For in the same year that Cronbach pronounced the death of the assessment center concept, Douglas Bray of AT&T engineered its rebirth and ultimate expansion.

The Management Progress Study

It has been said that the assessment center concept was adopted by American industry primarily as a direct result of this study. There is probably a great deal of truth in that speculation. Douglas Bray was hired by AT&T in the mid-fifties to institute programs of basic research on the progress of managerial careers. Bray conceived an ambitious longitudinal research design to identify and describe the nature of change or stability throughout the entire career of a manager. He hoped to ascertain the causes of changes, particularly those associated with salient company practices, and sought to identify the key predictors of managerial success. Though conceptualized as basic research in adult development, the study design included the use of an assessment center for tracking and predicting job progress. Therefore, the results of his research have led inevitably to conclusions regarding the validity of the assessment center concept for selection practices.

The Management Progress Study is still in operation today, some 23 years later. In 1974 a landmark text, Formative Years in Business: A Study of Managerial Lives, was published which describes in detail the results of eight years of follow-up interviews and reassessment, rounding out earlier publications which had described some individual aspects of the research (Bray, 1964; Bray & Grant, 1966; Grant, Katkovsky & Bray, 1967; Grant & Bray, 1969). Data for 16- and 20-year follow-ups are presently being collected and results will soon be published. For our purposes, the results presented in the 1974 text will be used for discussion.

The sample of subjects were entry-level recruits (all male) for management positions, both college-educated (n = 274) and not (n = 164). To obtain baseline data for the project, all subjects were put through a 3-1/2 day assessment center. *In contrast to nearly every other research project on assessment centers, the results for individuals on these assessments have never been shared with management*. Thus, the criterion (progress through management levels) has not been biased since management has not been able to use these data to make promotion decisions on individual cases.

The behavioral dimensions selected for measurement were derived from the literature, from other behavioral scientists and by executives within the Bell System. Bray himself "reconciled the many differences and settled on the 25 attributes" listed in Table 3 (Bray, Campbell, & Grant, 1974:18). Bray and his colleagues were in a situation somewhat similar to that faced by the OSS staff. They were not attempting to use the assessment center results to predict performance in specific

TABLE 3

Behavioral Dimensions of the Management Progress Study

Behavioral Dimensions	Definition
1. *Scholastic Aptitude*	General mental ability
2. *Communication Skills*	
a. *Oral*	How good would this man be in presenting an oral report to a small conference group on a subject he knew well?
b. *Written*	How good would this man be in composing a communicative and formally correct memorandum on a subject he knew well?
3. *Human Relations Skills*	How effectively can this man lead a group to accomplish a task without arousing hostility?
4. *Personal Impact*	How forceful and likeable an early impression does this man make?
5. *Perception of Threshold Social Cues*	How readily does this man perceive minimal cues in the behavior of others toward him?
6. *Creativity*	How likely is this man to solve a management problem in a novel way?
7. *Self-Objectivity*	How realistic a view does this man have of his own assets and liabilities, and how much insight does he have into his own motives?
8. *Social Objectivity*	How free is this man from prejudices against racial, ethnic, socio-economic, educational, and other kinds of groups?
9. *Behavior Flexibility*	How readily can this man, when motivated, modify his behavior to reach a goal?
10. *Need Approval of Superiors*	To what extent is this man emotionally dependent on authority figures?
11. *Need Approval of Peers*	To what extent is this man emotionally dependent on like and lower status associates?
12. *Inner Work Standards*	To what extent will this man want to do a good job even if a less good one is acceptable to his boss and others?

Table 3 cont.

Behavioral Dimensions	Definition
13. *Need Advancement*	To what extent does this man need to be promoted significantly earlier than his peers?
14. *Need Security*	To what extent does this man need a secure job?
15. *Goal Flexibility*	To what extent will this man be able to change his life goals in accordance with reality opportunities?
16. *Primacy of Work*	To what extent will this man find satisfactions from work more important than satisfactions from other areas of life?
17. *Bell System Value Orientation*	To what extent is this man likely to incorporate Bell System values, such as service, friendliness, and justice of company position on earnings, rates, and wages?
18. *Realism of Expectations*	To what extent do this man's expectations about his work life with the company conform to what is likely to be true?
19. *Tolerance of Uncertainty*	To what extent will this man's work performance stand up under uncertain or unstructured conditions?
20. *Ability to Delay Gratification*	To what extent will this man be able to work over long periods without great rewards in order to reach later rewards?
21. *Resistance to Stress*	To what extent will this man's work performance stand up in the face of personal stress?
22. *Range of Interests*	To what extent is this man interested in a variety of fields of activity, such as science, politics, sports, music, and art?
23. *Energy*	How continuously can this man sustain a high level of work activity?
24. *Organization and Planning*	How effectively can this man organize his work, and how well does he plan ahead?
25. *Decision-Making*	How ready is this man to make decisions, and how good are the decisions he makes?

managerial jobs. Rather, they were trying to describe the individual personality and predict progress through a series of management levels. Consequently, the list of "attributes" contained in Table 3 describes a mixture of intellectual abilities, attitudes, behaviors, skills and underlying needs. Many of these dimensions have been adopted for use in other present-day assessment centers.

The assessment center itself was composed of several tests of mental ability (e.g., School and College Ability Test, Contemporary Affairs Test), personality and attitude questionnaires (e.g., Edwards Personal Preference Schedule, Guilford-Martin), projective tests (e.g., TAT), and interviews. The standardized performance situations included a business game (in which assessees were required to create a small business, manufacturing simple toys, while being observed by two assessors); a leaderless group discussion; and an in-basket exercise (in which assessees were required to take some action on a series of 25 interrelated memos, letters and other items found in their in-basket in a simulated organization, and were then interviewed by an assessor regarding the reasons for the actions taken).

At the conclusion of the assessment center, the assessor team met for 1-1/2 days to review and rate each assessee's performance. All test scores and assessor reports on each assessee's behavior in the performance situations were read aloud and discussed. Then each assessor rated the assessee on each of the 25 dimensions using a 5-point scale. When there were serious disagreements over the ratings, further discussion was required until a consensus emerged. Finally, the assessors were asked to predict future success via a simple 2-point scale (i.e., would the assessee reach the third level of management within 10 years, or not?). Unfortunately, the exact manner in which this final judgment was reached is not fully articulated in the text describing the study.

Periodic interviews and other methods were used to follow the sample of asessees for 8 years, at which point they were reassessed, and their current management level recorded. A variety of analyses of these data has been published, but for this report we shall focus only on the results relevant to the assessment center predictions. Table 4 displays the correlations of each of the behavioral dimensions rated, with the criterion of management level at the 8-year reassessment. (Only 123 subjects were in this analysis due to attrition and use of the college-educated sample only.) The correlations obtained were, at their best, moderately low. Thus, ratings of these dimensions were not compelling predictors of upward mobility in the management hierarchy.

However, when the researchers compared the final predictions from the assessment center to actual management level attained, a more positive picture emerged:

TABLE 4

The Management Progress Study:

Correlations between Dimension Ratings and Management Level at 8-year Reassessment*

DIMENSION	LEVEL AT REASSESSMENT (N = 123)
Scholastic Aptitude	.19
Communication Skills	.33
Human Relations Skills	.32
Personal Impact	.15
Perception of Threshold Social Cues	.17
Creativity	.25
Self-Objectivity	.04
Social Objectivity	.13
Behavior Flexibility	.21
Need Approval of Superiors	-.14
Need Approval of Peers	-.17
Inner Work Standards	.21
Need Advancement	.31
Need Security	-.20
Goal Flexibility	-.18
Primacy of Work	.18
Bell System Value Orientation	-.02
Realism of Expectations	.08
Tolerance of Uncertainty	.30
Ability to Delay Gratification	-.19
Resistance to Stress	.31
Range of Interests	.23
Energy	.28
Organization and Planning	.28
Decision-Making	.18

* Adapted from Bray, Campbell, and Grant, 1974: 77

TABLE 5

The Management Progress Study

Assessment Center Predictions and Actual Management Level Eight Years Later*

Assessment Center Predictions		Actual at 8 years			
		3rd level	2nd level	1st level	Total
College sample	3rd level	64%	36%	-	100%
	not 3rd level	32%	68%	-	100%
Noncollege sample	3rd level	39%	61%	-	100%
	not 3rd level	9%	41%	50%	100%

* Adapted from Bray, 1976.

For the college sample, the assessors accurately identified 64 percent of those (remaining in the Bell System) who actually reached the third level of management, and 68 percent of those who did not. For the noncollege sample, the assessors were much better at predicting failure to reach the third level (91 percent) than success (39 percent).

Bray and Grant (1966) presented extensive analyses of other aspects of the research. They found that all of the measurement techniques used made some contribution to the final assessor judgments, though the performance situations exerted the most influence (and were also more predictive of progress). A combination of the behavioral dimensions was more predictive than any one dimension, and ratings of the individual dimensions were found to be highly related to the assessors' overall judgments of the assessee.

On the whole, the design, execution and analyses presented in the Management Progress Study are an impressive achievement. The results obtained to date have been fairly compelling to practitioners in establishing the validity of the assessment center concept. The research has also exerted a great impact on selection practices in the Bell System. To date, well over 100,000 Bell System employees have gone through some form of assessment center evaluation for selection or developmental purposes and quite a number of research studies have been conducted on these programs (to be reviewed in the fourth section of this report).

Because of the importance of this study as a cornerstone of the assessment center concept, certain aspects of the work require further comment. First, the researchers established the precedent of defining a large number of behavioral dimensions for measurement in the assessment center. Since the study was not targeted to a specific management position, the researchers were required to create dimensions which might be applicable to a variety of managerial jobs at the first three levels. These dimensions were developed, understandably, from available knowledge, practical experience and intuition. However, as a consequence, the dimensions were not equally specific with respect to the behaviors defining the dimensions. Thus, the assessors were required to use varying degrees of inference in rating an assessee on the dimensions. In short, the dimensions were a mixture of observable behaviors and underlying constructs of the personality. This study did little to clarify the distinctions between dimensions and behaviors. This is still a pressing issue in modern practice.

Second (and as a consequence of the first point), the research often failed to clarify how the dimensions were operationalized into ratings of assessee behavior by the assessors. The procedure utilized can be called into question. Assessors did not rate dimensions within each exercise (as had been done in the OSS center); rather, reports on each exercise were written, and all 25 dimensions were rated for each assessee at the end of the center. One might expect such a procedure to produce a considerable "halo effect"; indeed, the researchers found that the assessor's overall judgment of the assessee highly influenced his ratings of the dimensions. Thus, it is unclear whether the assessor ratings of the 25 dimensions were truly independent judgments of distinct aspects of the assessee. It is more likely that the assessors began to develop an underlying theory of the individual personality which led to ratings on the dimensions. This is a problem if the underlying theory is not made explicit, and raises the issue of what the assessment center process really measures--a point to which we shall return in the fourth section.

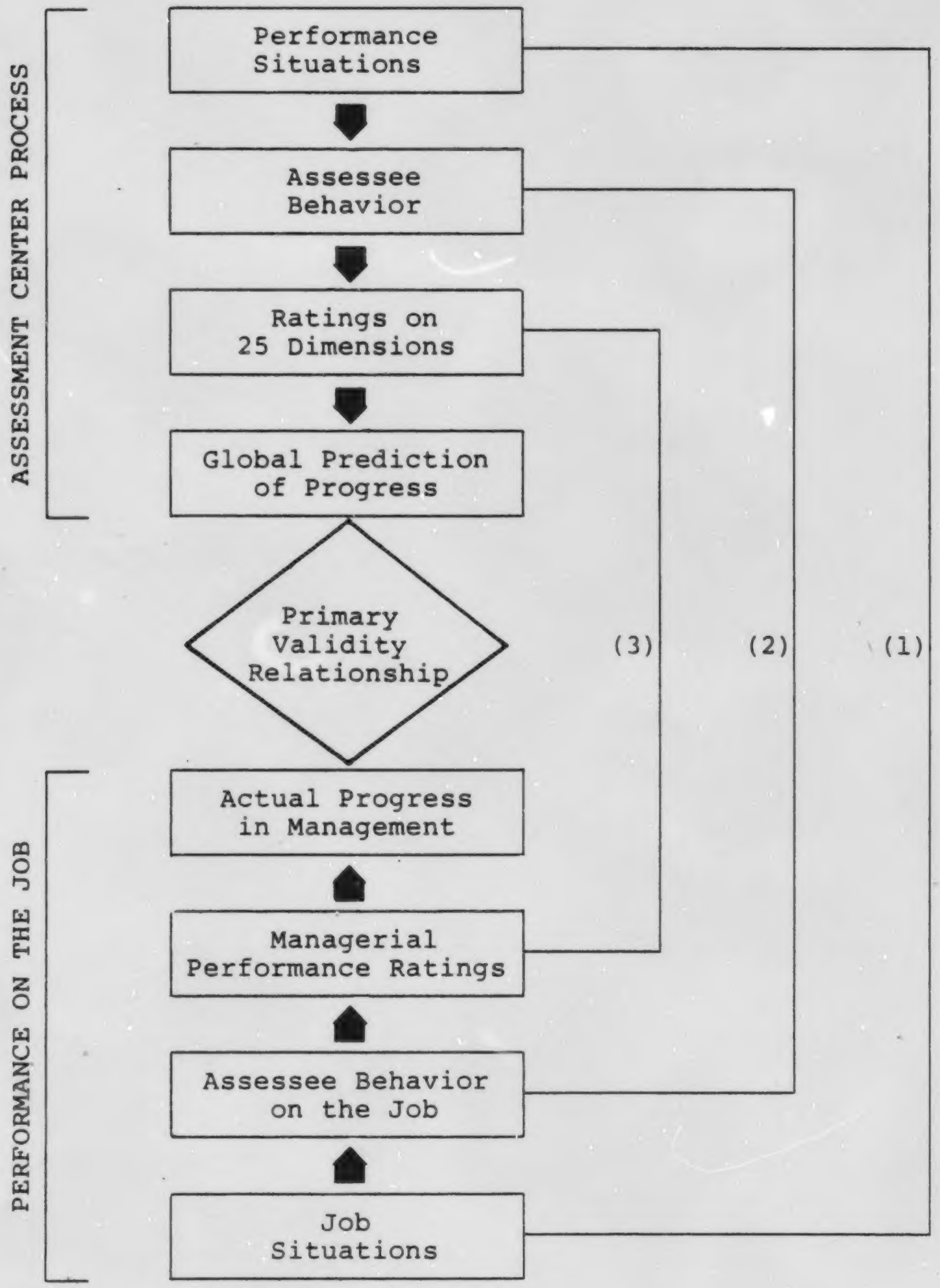
Third, the study established a research precedent which continues to appear in the literature. The 25 dimensions individually did not correlate strongly with the criterion of management progress, although clusters of the dimension were more strongly correlated (indicating possible halo effects as suggested above). The best predictor of the criterion was found to be the overall prediction at the end of the assessment center (i.e., will the assessee reach the third level of management or not?). Thus, the process of pooling ratings across the 25 dimensions, and across the assessor team, undoubtedly held the key to the predictive power of the assessment center process. Yet it is unclear how these multidimensional ratings were collapsed into this final overall judgment. The process required considerable data reduction by the assessors. In the absence of explicit rules, or a theory, to arrive at this overall judgment, what rules were actually used? Without further specification of the determinants of this judgment formation process, the relationship of numerous specific judgments to a final global prediction remains open to question.

Finally, the study established a clear precedent for the validity of the assessment center concept; therefore it is critical to understand the nature of the validation strategy used. This strategy, as outlined in schematic form in Table 6, continues to be the predominant paradigm in assessment center research. The primary validity relationship, at the center of this diagram, is the correlation of global assessment predictions of managerial success (in this case, progress) with a global measure of actual success in the organization (i.e., upward mobility in the hierarchy of the organization). Underlying each side of this validity equation were a series of more specific situations, behaviors, and judgments with some presumed equivalence (as indicated in the diagrams by lines 1,2, and 3). Thus, the logic continues, the discovery of strong correlations between these global measures lends credence to the validity of each step of the assessment center process. Therein lie some difficulties.

A basic tenet of the assessment center concept is that the performance situations, and assessee behavior stimulated, are representative of actual job situations and assessee behavior on the job. This is, in essence, the idea of content validity. The Management Progress Study could not, and did not,

TABLE 6

Schematic of the Validation Strategy of the Management Progress Study


ASSESSMENT CENTER PROCESS
Performance Situations
Assessee Behavior
Ratings on 25 Dimensions
Global Prediction of Progress
Primary Validity Relationship
(3)
(2)
(1)
PERFORMANCE ON THE JOB
Actual Progress in Management
Managerial Performance Ratings
Assessee Behavior on the Job
Job Situations

attempt to concretely assess content validity of its assessment techniques. Today, however, with an increasing emphasis being placed on content validation of tests, the automatic assumption of content validity without some rational systematic evidence is no longer acceptable.

A more complex assumption in this logic is that the assessor ratings of the behavioral dimensions anticipate how AT&T managers on the job, viewing the assessee's performance, will rate that job performance. Since Bell System managers did not use the same 25 dimensions in their evaluation of the assessee's job performance, one must assume that the assessment center dimensions did in fact measure something which the manager took into account in assessing job performance for promotion. This is, in essence, the concept of construct validity. This study (and most studies conducted since) did not really correlate behavior ratings in the assessment center with behavior ratings on the job, but rather examined the relationship between global judgments. Thus, one cannot, with certainty, establish what the assessment center actually measured with respect to on-the-job performance. What one can say is that global predictions of success, however reached in the assessment center, correlate with global measures of success on the job. Why this relationship obtains can be subject to a number of interpretations, and is a subject of some debate in the current literature.

In summary, the Management Progress Study made a substantial contribution to the emergence of the assessment center concept in practice and largely overshadowed the debates, within academic circles, concerning the viability of the concept. Many of these debates have begun to resurface in recent years, though practical applications of the assessment center concept have continued to expand at a rapid pace.

The Coming of Age of the Assessment Center Concept (1960-1979)

The decade of the 1960's witnessed the slow diffusion of the assessment center concept into practice within AT&T, as centers were created for the selection of first-line supervisors, salesmen, and middle-level managers in a number of the operating companies. Diffusion into other corporations occurred at a somewhat slower pace.

Word of the AT&T work filtered through informal and formal organizations of industrial psychologists, resulting in several applications in a few large corporations (the most important of which were IBM, SOHIO, Sears, General Electric, J.C. Penney, and Wickes). The majority of these applications appeared to be designed for selection into first-level supervisory positions and/or middle-level managerial postions, though some centers were beginning to be utilized for the early identification of managerial potential and as part of training programs in managerial skills.

By 1965, several longitudinal research programs had been initiated in these corporations (particularly IBM, Sears and

SOHIO) to validate further the assessment center concept. Results of the Management Progress Study began to be published, along with results from research efforts underway elsewhere. In addition, information regarding these applications began to be exchanged at symposia and conferences on the technology. Nevertheless, by 1970, only some 12-15 organizations were widely known for their applications of the assessment center concept.

The onset of the 1970's brought what can only be described as an explosion in the diffusion of the assessment center concept into practice throughout the United States and in several other countries (e.g., Norway, England, South Africa, Canada, Germany and Italy). Applications began to appear, under a variety of labels, in somewhat smaller organizations in the private sector as well as in a number of federal, state, and educational settings. The objectives of the centers also began to expand. While many were still designed primarily for selection and promotion purposes, quite a number were created for early identification of potential, career development for employees or as part of training and development programs. The majority of these applications continued to reflect organizational concerns for selection into first-level supervisory positions and middle-management positions, though several applications reflected a concern for long-range development of employees for various general management jobs. Novel applications also began to appear for selection into sales or technical positions, as well as for selection of personnel for work teams in plants.

By the end of the decade, as estimated by various writers, somewhere between 500 and 2,000 assessment centers were in operation in the United States. While these estimates are difficult to verify, it is apparent that the growth rate in applications of the assessment center concept has reached the proportions of a general social movement, perhaps equalled only by the small-group fads (encounter groups, T-groups) of the 1960's. As a consequence, the components of the assessment technology have undergone considerable elaboration in practice. A plethora of behaviorial dimensions have been defined for various purposes. Some new performance situations have been created to simulate specific job demands, though many centers continue to rely upon leaderless group discussions and in-baskets as core techniques. A variety of training programs have been instituted for assessors, who serve the centers for varying periods of time. The manner in which final competency judgments are rendered and reported back to the assessee (or the organization) has become a widely varying practice. Many organizations have begun to institutionalize components of the assessment center technology as part of standard personnel procedures (e.g., performance appraisal or compensation review).

The details of this current practice have been reserved for the next section of this report and will not be further elaborated here. It is sufficient to note that the last ten years has brought the coming of age of the assessment center concept. However, one central question requires further comment: given the somewhat lengthy history of the assessment center concept, why the relatively sudden outburst of applications in the last decade? We believe several key factors underlie this phenomenon.

The first factor is the rather powerful appeal of the assessment center concept itself to managerial audiences, particularly to human resource development specialists. The last 20 years have brought a dramatic increase in the application of behavioral science theory and technology to management practice. Testimony to this phenomenon were the birth of such disciplines as organizational development and management development, the greater resources being applied to training, and the creation of human resource development departments within many organizations. As a result, a multitude of measurement techniques have been devised for diagnosing, or assessing, human behavior in organizations. Far too many of these techniques have lacked a substantial theoretical or research base, and some are viewed by managers as possessing questionable utility for improving the management function.

The assessment center concept presents a rather sharp contrast. Compared to other techniques, the assessment center concept is rooted much more firmly in a tradition of research. Many of the earliest centers were staffed by a mixture of psychologists and managers, and insistence on measurement rigor was clearly present. (This is, perhaps, less true today with the rapid expansion of applications, but the aura of scientific rigor is still a significant component of the appeal of assessment centers.) Practitioners have also actively promoted the image of the assessment center as a scientifically rigorous measurement technique. Given the competition of other, less-researched techniques, the assessment center concept is an attractive alternative.

Yet this is only half of the picture. It is our experience that managers often are attracted to the concept because it *looks* valid (i.e., it has "face validity"). At first glance, the components of the assessment center technology make a great deal of sense to managers. Every manager has a theory of managerial success which can be reflected in the behavioral dimensions. Observing job candidates in job simulations provides much more direct evidence of ability than simple face-to-face interviews. Using several assessors to make a selection judgment seems much better than using one judge, and one can still exercise one's managerial expertise and experience. Despite its psychological research base, one does not have to be a psychologist to assess the job candidate. In effect, the assessment center concept--at least on the surface--does not represent a radical departure from the use of good, old-fashioned managerial common-sense!

Thus, the appeal of the assessment center concept can be viewed as originating in the combination of two strong, but contrasting, images--rigorous scientific measurement and managerial common sense. While persuasive to managerial audiences, such imagery is a source of concern to many practitioners and writers in the assessment center field, and was likely an underlying factor in the development of standards for practice. Regardless of the reality beneath the imagery, the rapid growth of assessment center applications has been described as bordering on religious fervor and tainted with a "faddish commercialism" (Finkle, 1976). This is perhaps inevitable in our culture, but is likely to represent a serious constraint on further research and development of the assessment center concept, even though the number of applications is likely to continue to increase.

Certainly a second key factor spurring the growth of applications is the voluminous literature appearing, in the last ten years, on the concept. A few more research reports have been published, while numerous "how-to-do-it" articles have been published in a variety of professional and trade association journals. William C. Byham, the creator of the J.C. Penney assessment center and co-founder of Development Dimensions International (the largest supplier of assessment center materials and consulting expertise in the United States) established the tone of these articles in a seminal review of the assessment center concept for the Harvard Business Review in 1970. He summarized:

> Reports have proved to be remarkably valid. Longitudinal studies of thousands of employees assessed over the last

> few years indicate that *this assessment method is much more accurate than traditional appraisal procedures*, and these seem to be the reasons:
>
> - The exercises used are designed to bring out the specific skills and aptitudes needed in the position. . .
> - Since the exercises are standardized, assessors . . are able to make valid comparative judgments.
> - The assessors . . . being emotionally disengaged . . . are unbiased.
> - The assessors . . . can pay full attention to the candidates' behavior in the exercises.
> - The procedures focus their attention on the primary kinds of behavior they ought to observe in evaluating a promotion candidate.
> - They have been trained to observe and evaluate these kinds of behavior (Byham, 1970:151).

Such a refrain has been echoed in various publications, by lesser-known authors, throughout the decade.

In addition, nearly every year since 1972 has brought the publication of a major review of the assessment center concept and associated research. A perusal of these reviews is rather fascinating. Most of the authors share at least a guarded optimism for the concept--some are openly enthusiastic advocates--but many caution the reader that further research is necessary on several fronts in order to disentangle the fundamental determinants of the validity of the assessment center concept. Let's briefly examine a few of these authors' comments.

Fleming and Sells (1972) suggested that the assessment center is a difficult and complex process that has not been placed into practice under consistently ideal conditions. While finding the validity studies (to be reviewed in the fourth section of this report) to show moderately high correlations, they questioned whether these results were greatly superior to other measurement techniques. They explicitly noted that their interpretation of the research studies was in apparent contrast to those made by several better-known practitioners.

Kraut (1972) argued that his own research at IBM and other studies, while possessing methodological flaws, showed a consistent pattern of apparent validity. He pointed to the need for more research on the impact of assessment centers on organizations, assessees and assessors. He further identified the need to isolate more specifically the human characteristics measured by the assessment center process, and echoed the concern for rigorous standards in operating centers.

Wilson and Tatge (1973) shared Kraut's concern over what assessment centers really measure, and presented some data suggesting that the process predominantly measures skill in interpersonal relations. They further argued that "there is simply no available or published evidence to support claims of the significant superiority of the centers over other methods" (1973:179), and cited little evidence to substantiate the claims that the process is more cost-effective or provides ancillary benefits for assessors or assessees (in terms of training or development).

Bender (1973) conducted a limited survey of practice resulting in cautions regarding the misuse of assessment center data by line managers and the need for more consistent training of assessors in observation techniques.

MacKinnon (1975), one of the original workers in the OSS program, concluded that the research evidence, while confined to only a few programs, did demonstrate higher validity for assessment centers than traditional paper-and-pencil measures. He felt final judgment should be reserved until more research was conducted, particularly research aimed at accounting for why the process might be more valid.

Finkle (1976) suggested the volume of research was encouraging, though limited in terms of the number of programs actually researched. He argued that, with the expansion of practical applications, the resulting variations in practice have muddied the generic label of "assessment center," and cited a critical need for close professional guidance and supervision to avoid the "faddish" implementation of the concept.

Finally, several authors have provided substantially positive reviews of the assessment center concept, though most of these authors are widely known as advocates of the method (Bray, 1976; Cohen, Moses & Byham, 1977; Moses & Byham, 1977; Huck, 1973). Byham (1978) has published a more recent article in which he has attempted to identify key factors associated with the failure of assessment centers to reach their full validity. He isolates several sources of error:

- Failing to conduct thorough and up-to-date job analyses for the creation of behavioral dimensions;

- Failing to design performance situations which truly sample job situations, and recreate the specific demands of the target job;
- Failing to adequately train assessors in the observation and rating of behavior;
- Failing to control the assessor team discussions, e.g., by forcing ratings to be backed up by concrete behavioral observations, or by disallowing ratings to be based on a comparison with other assesses; and
- Failing to specify the purpose of the assessment center in the organization's decision making, and failing to inform participants of the use of the competency judgments.

In summary, the growth in the published literature during the last decade is a second key factor underlying the growth of applications, and has brought the concept into greater visibility. This increased visibility has also engendered some debate, regarding many of the components of the technology, between the enthusiastic advocates and the somewhat more cautious skeptics. We feel that such debate is healthy, and will surely continue in the next decade.

A third factor which clearly contributes to the increased use of assessment centers is the variety of governmental, judicial, and legislative directives concerning civil rights and equal employment opportunities. The use of psychological tests for employment decisions has come under heavy attack from both within and outside of the psychological community--particularly the use of the more traditional, academic, paper-and-pencil tests. The results of myriad court decisions have identified the key to acceptability of test procedures to lie in the "job-relatedness" of the test (i.e., its content validity). Given the emphasis of the assessment center concept on the observation of behavior in job simulations, many organizations have implemented the method to avoid possible legal action against their selection procedures. In at least two instances, an assessment center was designed to replace a prior system as the result of legal dispute (at AT&T and at the EEOC itself!). In short, the technology and concept of the assessment center has placed it in a relatively strong position vis-à-vis the courts.

It should be emphasized that the mere implementation of an assessment center does not guarantee that it will be legally defensible. Richard Barrett, a well-known psychologist who has testified in over 70 EEOC cases and who played a large role in the creation of the most recent EEOC guidelines regarding the use of tests in employment decisions, suggested in the 1979 International Congress on the Assessment Center Method that assessment centers do appear more defensible; yet poor quality in the design of any components of the technology can easily be attacked. Thus, with proper caution and rigor in the design of the center, organizations might greatly reduce their risk by application of the concept. It would appear that many companies have been greatly influenced by this type of statement.

A fourth factor in the growth of applications can be seen in the proliferation, primarily since 1970, of consulting firms and other advocacy groups. Two of the largest such firms

(Development Dimensions International [DDI] and Assessment Designs, Inc. [ADI]) were founded by well-known writers and researchers (Bray and Byham, and Jaffee, respectively). These two firms, as well as a few others much smaller in scope (e.g., American Management Association, Educational Testing Service, Educational Research Corporation, and others), have been largely responsible for the development of numerous standardized packages and materials for assessment center design and assessor training. They also provide training programs for assessment center designers, and offer on-going consulting expertise to client organizations implementing the concept. As a consequence, these firms have reduced the start-up costs for the smaller company unable to develop its own center, and have pioneered innovative applications to a variety of selection, development and training needs.

Both of the larger firms have also instituted regular publications regarding the assessment center concept. DDI publishes Assessment and Development as a regular newsletter detailing key events in the assessment center field, and ADI publishes a quarterly journal for research articles entitled the Journal of Assessment Center Technology. DDI was also instrumental in the founding, in 1973, of the International Congress on the Assessment Center Method. This meeting is held yearly for the exchange of research and practical ideas among user organizations, practitioners, and researchers. The most recent meeting, held in June of 1979, was attended by well over 100 participants.

Principals from both of these firms, and many other well-known researchers, founded the Assessment Center Research Group in the early 1970's. This group is primarily responsible for a recent, and quite comprehensive, text on the assessment center concept, *Applying the Assessment Center Method*, by Moses and Byham, 1977. In conjunction with the International Congress, many of these individuals are responsible for the creation of the standards contained in the document cited in the introduction of this report.

The creation of these consulting firms, and the relatively tight network of researchers and practitioners in associated groups, has contributed substantially to the growth in application. It is quite likely that these groups will continue to exert a great impact on the development of the concept for years to come.

In summary, it is fair to say that the last 20 years have brought the assessment center concept to the forefront of the more general field of competency assessment. The concept is becoming rapidly diffused to a broad array of organizations, and the field has gained a great deal of momentum. What then can we say that this history has taught us?

The Lessons of History

The origins of the assessment center concept can be traced over some 40 years both to the practical demands of large bureaucracies for new selection systems to combat critical manpower problems, and also to the more theoretical interests of psychologists in designing more effective measurement tools to describe the individual personality. These dual roots of practical problem solving and theory-based research have heavily influenced the evolution of the concept, and clearly form the basis for the phenomenal popularity of the concept in current practice.

The interface of research and practice can also be a significant source of tension and debate among researchers and practitioners. Our examination of the history of the assessment center concept would suggest that the demands of practice have often raced ahead of the development of a supporting research base (though such research has certainly never been totally lacking), leading researchers to raise important questions regarding the true extent of data-based knowledge upon which common practices are based. With the rapid growth in the present number of applications, one could reasonably argue that this interface of research and practice is in pressing need of reintegration and rearticularion. Such developments as a set of standards for the field, the publication of research journals, and the publication of general texts are undoubtedly steps in the right direction; however, greater collaboration between researchers and practitioners is needed in order to pinpoint the research base for the design of many components of the assessment center technology.

Of the many questions which have been raised over the past years, at least two general questions are of paramount importance. First, throughout the evolution of the assessment center concept, the question of what the assessment center process really measures has been continually raised. Since the design of the OSS program and AT&T's Management Progress Study, the creation of a set of dimensions for measurement has become a standard component of assessment center technolgy. Murray and the OSS staff viewed these dimensions as generally representative of underlying constructs of the human personality as evidenced in behavior on tests or in simulations. As a result of post-war applications, the practice of inferring underlying personality constructs was resoundly attacked and discredited as a procedure open to serious error. The Management Progress Study, which was highly influenced in design by the OSS program, continued the use of some underlying personality constructs, but also started the trend towards definition of dimensions as aggregates of observable behavior. Today, the standards for practice prescribe the use of some type of job analysis to identify dimensions for measurement. As a consequence, these dimensions are often held to describe the behavioral demands of the job, though in practice the dimensions are often just as interpretable as representing underlying capacities of the individual to perform the job. What, then, is the true nature of a behavioral dimension? This becomes an even greater issue in situations, such as in educational settings, where one is attempting to assess the basic competence of an individual to perform in a variety of situations. Is the assessment center attempting to measure behavior or the underlying capacities of the individual--as represented in the notion of competence? This is a conceptually thorny issue which needs to be addressed by further research and theory.

This issue has clear practical implications for the definition of the dimensions per se, as well as for the translation of dimensions into observations and ratings of behavior by the assessor staff. From the Management Progress Study forward, the research base has suggested that assessors do not really treat a large number of behavioral dimensions as independent measures. It may well be the nature of our information-processing capacities to reduce the complexity of another person's behavior to a few critical underlying attributes. This is not to suggest that assessors should not be forced to anchor such judgments in observable behavior; however, we would suggest that the average assessor may not be able to accurately discriminate ratings on a great number of dimensions. As a consequence, it is not surpising that individual dimensions do not often correlate highly with criterion measures, or that clusters of dimensions are better predictors; indeed, we would expect this finding. What is critical is determining how, and what, the assessor is actually rating via the use of a series of dimensions, making these inferences as explicit as possible, and measuring these same constructs in the assessee's on-the-job performance. Therein lies the answer to the question of what the assessment center process truly measures.

The second research question arising from the evolution of the assessment center concept is, to some extent, a more general statement of the previous issue. Since the first applications of the concept, researchers and practitioners have been quick to address the question of "are assessment centers valid?" This research has often been driven by the practical need to justify the costs of implementing the concept for selection and development purposes (in addition, of course, to the research sensibilities of practitioners). A good deal of research has been devoted to this question, and as many reviewers of this research have commented, there is a fairly consistent pattern of validity--primarily in the relationship of global predictions of competence as assessed in the center with global measures of success on the job.

With the growth in applications, a great opportunity is available for the conduct of research on an even more fundamental question, namely, "why are assessment centers valid, and what aspects of the technology contribute to, or hinder, its capacity for valid predictions of competence?" It would appear that at least some researchers and practitioners are now beginning to approach the concept from this perspective (for example, there would appear to be less concern today with establishing the superiority of the assessment center method over other appraisal techniques, and more concern with the unique contribution of various types of techniques--i.e., interviews, simulations and tests--to overall competency judgments). Nevertheless, it is important to emphasize that research devoted to answering these questions would represent some departure from current practice. Such research would also address many issues that have been raised since the early 1950's regarding components of the assessment center technology, such as:

- What aspects of performance situations are critical to eliciting the assessee behavior which is deemed to be measured in the situation;
- What characteristics of the individual are best measured in which types of performance situations or tests--and why;
- What is the best format for the design of assessor training for the observation, recording and rating of behavior;
- What are the key sources of error in behavioral observation and rating;
- What variables impact the dynamics of assessor team discussions, and how can these be best controlled; and
- What is the best means for summarizing multiple ratings of behavioral dimensions into one overall competency judgment?

These questions are but the tip of the iceberg. The assessment center concept today presents an intriguing challenge for both researcher and practitioner.

In effect, perhaps the greatest lesson to be derived from the history of the assessment center concept is that practice will most likely continue to grow--at least for the present--regardless of the research base. This represents both a challenging opportunity for real advancement in assessment technology and also a serious threat to the underlying complexity and sophistication inherent in the concept. The continued integration of research and practice will require, in the coming decade, a greater devotion to the development of a general theory of the assessment center process--a theory which integrates past research, identifies the key research questions for the future, and provides data-based standards for future practice.

Because of the complexity of the assessment center process, such a framework will, of necessity, cross many different disciplines of psychology, and should draw upon the ample research available in various fields to address the questions we have raised in the discussion above. In the absence of such an integrative framework, many of the key questions concerning the current state of the practice will continue to go unanswered, and the state of the art will not be significantly advanced.

This argument for the consistent integration of research, practice and theory is even more important in the context of the relatively new extension of the assessment center concept to educational settings. Even though such applications are already under way, very little published literature is available on these programs. It is likely that these experiences will evolve slightly different practices and procedures. The availability of a more general theoretical framework could well ease the adaptation process and perhaps signal the critical areas of concern for this work. In short, we would agree with the assessment of Kurt Lewin, one of the progenitors of modern social psychology, that there is nothing as practical as a good theory. We will return to this theme once again in the final section of this report as we discuss the further implications of the state of the art for application to education.

III. PRESENT STATE OF THE ART IN PRACTICE

Nature of the Sample

As part of the mandate from NIE, we have attempted to contact as many assessment center users, suppliers, and designers as was possible. Before detailing the nature of this sample, it is important to recongize the process by which it was generated. Under the conditions of our contract with NIE, we could not conduct a mail survey to identify the total universe of assessment center users, practitioners, or designers. Therefore, we resorted to a "snowball" technique whereby we identified assessment center users through the professional literature, through the management literature, and through personal contacts of the research team in education, industry, and consulting organizations. This method has both advantages and disadvantages. A key advantage is that where we had direct contact with an individual in an organization using the assessment center method, we were more apt to get candid and detailed information on the actual operation of their center. On the other hand, the disadvantages are also clear. Our sample is not a rigorous random sampling of all users. We followed a network of personal and professional contacts; therefore, our web of contacts spread out in a way different from the contact network another research team would have evolved using the identical method. We were fortunate, however, to have participated in the 1979 International Congress on the Assessment Center

Method, at which we were able to make acquaintance with and pursue the contacts and networks of other professionals versed in the field. Those contacts helped us to identify other users and expand our network beyond that which we had already developed.

The major suppliers of assessment center materials have estimated that there are upwards of 1,000 assessment centers used in this country today. While we found it difficult to verify, or even identify, all of these users, it is possible that such a rough estimate is reasonably accurate. Given that there may be 1,000 or more assessment center users today, we have identified through our sampling procedure about one-quarter of those users. By "identification" we mean that we know the names of user organizations and, in many cases, the consultant suppliers with whom they work. We attempted to contact as many as possible of this group of users. As a result, 53 organizations were willing to talk to us about the use of assessment center methods in their organization.

The nature of the information which those organizations were willing to share is of considerably varied quality and detail. The assessment center method is very popular with management in a variety of industrial and educational settings. For that reason, most organizations were willing to share with us, at the very least, their enthusiasm and commitment to the method. They were somewhat less likely to share detailed research or design information with us. In some cases, that hesitancy reflected a concern for proprietary information; in other instances, it reflected a concern about EEO issues, or making the validity of selection and promotion systems available for public scrutiny. Nonetheless, it is important to note that, in general, testing and measurement professionals were far less reticent to discuss their assessment centers than they were to discuss other selection practices. We would speculate that their reaction may well reflect concerns regarding the perceived legal vulnerability of traditional selection methods, such as paper-and-pencil techniques, versus the more face-valid assessment center technology.

A more detailed description of this sample can be found in Tables 7A, 7B, and 7C. (For ease in reading, all tables on the current state of practice have been placed at the end of this section of the report.) Table 7A lists, by name, all users and designers that we could identify. Table 7B groups these organizations into 13 broad categories, and describes the quality of data we were able to obtain from our contacts. Industrial applications account for well over 55 percent of our sample, government applications for nearly 25 percent, and educational applications for just under 15 percent. Finally, Table 7C lists the job titles to which these applications are targeted for those applications where this information was attainable.

It is not unreasonable to argue that an interview sample of slightly more than 50 assessment center users is a limited sample of the universe of 1,000 or more actual users of the technology. However, there are arguments to be made for the representativeness of this particular sample. As we have reported previously, the assessment center concept has enjoyed burgeoning popularity during the past two decades. During that time, a number of consultants and consulting organizations have entered the business of providing materials and services to organizations which want to implement assessment centers. Some individual consultants and organizations have successfully implemented assessment centers of their own, and have done some very innovative work in the field; nevertheless, they are in the minority. There are, in fact, a small number of assessment center suppliers who service the vast majority of assessment center users. We have discovered through our interviews with assessment center users that the names of two or three suppliers are repeatedly mentioned. Thus, there appears to be considerable overlap in the method, in the materials, and in the rationale and operation of most of the assessment centers which we have identified. By virtue of this fact, we are much more confident that our limited sample is fairly representative of the current state of practice in assessment center technology.

We also believe that we have adequately sampled the educational use of assessment center methods. While educational settings are less common as a backdrop for the use of assessment centers, there are still quite a number of applications of the method in education today. We found that those applications are less likely to originate from the major suppliers of industrial applications. Also, the educational users were less likely to be constrained by business competition, and were more likely to know about, and share with us, the names of other users in educational settings. We have tried to identify more completely the sample of educational applications, though we did not generally have the luxury of using client lists from organizations which exclusively supply educational assessment centers. One small organization which does supply assessment centers to education has, in fact, shared its list of clients with us; however, this list is far more limited in number and in scope that are the lists which come from the industrial suppliers.

In summary, we are confident that the sample is highly representative of typical applications in the educational arena, and fairly representative of the range of applications in industrial and governmental settings. From these contacts, we have attempted to identify what is truly typical of modern practice with respect to a number of salient issues, which we now examine in detail.

Reasons for Setting Up An Assessment Center

The first question that comes to mind when one is trying to identify the use and users of assessment centers is why an organization chooses to set up an assessment center at all. In response to this question, organizations provided us with a large number of salient reasons. Their comments are presented in Table 8. (Please note that this and all the following tables detailing the state of practice present the data which were of sufficient detail to include. Thus, uneven numbers across tables merely reflect variability in how candid organizations were around the various topics.) Table 8 suggests that there are a wide variety of reasons given by organizations for choosing to use the assessment center method. The most common themes among the reasons cited are the following: (1) improving employment and personnel selection; (2) minority involvement; EEO issues and affirmative action; (3) a need to evaluate the skills of individuals in nontraditional settings; (4) chronic turnover problems in the organization; and (5) face validity of the selection method. There are a number of other equally reasonable reasons for instituting the use of an assessment center; the reader should consult Table 8 for greater detail. However, let's examine some of these more common themes in order to understand the predominant rationale for the use of assessment centers in industry and education.

First, a number of organizations said that they were concerned about improving employment and personnel selection practices. Industry and education have become disillusioned with the standard paper-and-pencil testing methods of the recent past. Academic research and judicial decisions have driven home the point that a paper-and-pencil test, or a single manager's evaluation, may not reliably and validly tap the important skills that an individual possesses for a job. The assessment center is clearly viewed as a significant advance over such traditional procedures.

A second reason often cited for implementing an assessment center was a concern for involving minorities in the organization, taking affirmative action, and an overall concern with EEO issues (both legal vulnerability and ethical concerns). It is not surprising that this area of concern is important to business organizations today. Industries have become increasingly aware of the potential deficiencies in their own selection and testing practices. This awareness has been prompted greatly by tests of the practice which have come before the courts. The courts have in the recent past shown increasing concern with the validity of selection devices, with the potential adverse impact of selection devices, and with the reasonableness or face validity of those selection practices. While the research is certainly not definitive at this point, there is at least some evidence that assessment centers are reasonably valid, are reasonably well accepted by participants in terms of their face validity, and can be reasonably expected not to have adverse impact on protected classes. (See Section IV for a further discussion of the research foundations of this faith in the assessment center concept.)

The third reason often cited for implementing an assessment center was the need to evaluate skills of individuals in nontraditional settings. This rationale comes most often from educational applications of the assessment center method; however, it is a rationale that is also related to the needs of industry. There are a number of jobs in industry and a number of positions in education for which there is no reasonably valid method for certifying competence, whether by some paper-and-pencil measure or other traditional means. Education and industry alike have found that the assessment center may be more appropriate, and richer in content, for the assessment, identification, or evaluation of critical skills. Furthermore, the assessment center is viewed as more meaningful in helping individuals make important career decisions based on these diagnoses.

The fourth and fifth rationales are closely related. A number of organizations (e.g., the insurance industry) have noted that turnover in an organization is very expensive and is also very common among new hires. Therefore, they have turned to the assessment center method because they feel that it is both a better method for assessing skills needed on the job, and a better way to give the applicant a realistic preview of what he or she will confront on the job. The rationale here is that by more rigorously assessing applicant skills, those most apt to stay, and succeed, in the job will be hired. The realistic presentation of job information to a candidate is intended to prevent turnover by preparing an employee to cope with both the positive and negative aspects of the job. Face validity is also an important consideration in assessment center use because the performance situations not only provide new applicants with realistic expectations about the job, but also provide job incumbents who are up for promotion or are attending developmental centers a sense that the process is meaningfully related to the job. This face validity also has incidental benefits for EEO issues. Selection and testing methods which appear on their face to be related to the job seem less likely to engender employee resentment or to lead to suits brought against the organization by an unsuccessful assessee.

Resources Required for Assessment Center Design

We were struck by the pervasiveness of a limited number of assessment center suppliers when users identified the resources they employed to design and implement their assessment centers (see Table 9). This table suggests that there are several organizations which did all of the design and implementation work for their own assessment center. However, there are three times as many organizations which relied on external consultants. The consultants who are most prominently named, and whose materials are most pervasively used, are Development Dimensions International, Assessment Designs, Inc., and the corporate research group at AT&T. In addition to these three major suppliers, materials and assistance for design were provided to organizations from groups at a variety of universities, from a small consulting group called Educational Research Corporation, and other groups such as Harbridge House, Personnel Decisions, Inc., the Center for Creative Leadership, American Management Associations, and the counsulting firm of Zenger and Miller (formerly with Syntex). In virtually all of the organizations which did use external consultants, in-house personnel, staff, or psychologists were enlisted to support the efforts of the consultants. This support most often took the form of (1) doing the job analysis for the consultant prior to design of instruments; (2) continuing the process of assessor training; or (3) administering the center after the initial pilot test.

Purposes of the Assessment Center

Organizations may have many different reasons for implementing an assessment center; however, once the assessment center is in place, there would appear to be a limited number of purposes to which the assessment center data can be applied. These purposes essentially reduce to using the assessment center for one of three activities: (1) to select people for a specific job; (2) to promote people within a career path within the organization; or (3) to use the assessment center data for constructing a career development plan for an individual. Despite the fact that most organizations could describe the global purpose of their assessment centers, few could articulate concrete criteria associated with the purpose; that is, few identified specifically what they were trying to predict, measure, evaluate, or anticipate with the assessment center data. Whether they had, in fact, never clearly stated measurable criteria, or were just unwilling to tell us how they measured good performance in their organization, is somewhat unclear.

An examination of Table 10 would, at first glance, suggest that selection and development are equally common purposes cited for an assessment center. However, few organizations used the term "promotion" to describe selection of internal candidates for new jobs. Therefore, if we combine the references to both selection and promotion, we find that more personnel decisions concerning employment status are made on the basis of assessment center data than are personnel decisions concerning employee career path development. Thus, while there are certainly more developmentally-oriented centers in operation today than in the past, the more common purpose of assessment center application would appear to be for employment decisions.

In addition to identifying a number of organizations that were already using an assessment center for selection, promotion, or developmental purposes, we also identified a number which were at an early stage of designing and piloting a center for use within the organization. Many organizations that were at this stage of development were experimenting with methodology and had not given great concern either to the ultimate purpose of the center or to the use of the data from the center. It seems reasonable that in the process of designing an assessment center, one should keep its purpose in mind. Nonetheless, this foresight seems to be lacking in many of the pilot studies we encountered. For example, in one organization a center was being piloted as a developmental program because the organization recognized that they could not take an untested technique and use it for selection or promotion decisions. However, the ultimate purpose of this center was to be a component of the selection process. Clearly, though perhaps not so clearly to the user, an assessment program which is piloted developmentally will have different effects and problems than will those same assessment exercises implemented for purposes of making selection decisions. This issue seems to be overlooked in some organizations.

A central concern of any assessment center operation is the validity of the judgments and evaluations made in the center. In the sample that we contacted, we discovered a discouraging, though perhaps not surprising, trend in the practice of validating assessment centers. As indicated earlier, assessment centers have often been adopted because of their apparent face validity and their apparent defensibility vis-à-vis the courts. This perception may well be resulting in far less concern with attempts to establish the true predictive validity of the assessment center process, by many practitioners in organizational settings. Of the 53 assessment centers for which we have data, 16 report criterion validation of their assessment center, and 4 more report content validation. Twenty-two, however, report that the assessment center has never been validated, and only 9 (including pilots) report that they are currently collecting data for purposes of validation.

Some issues concerning the methods and importance of validating assessment centers should be mentioned at this point. First, there are a number of assessment centers which are operated for developmental purposes. It was not uncommon among our sample to find assessment center administrators who claimed that because the assessment center that they were running was developmental, there was no need for them to validate the center. Strictly speaking, in terms of the legal defensibility of an assessment center, developmental assessment centers need not be validated. The courts are empowered to sit in judgment on the job-relatedness of devices used as the basis of employment decisions; they are not empowered to oversee counseling, training, and career advice such as might be the result of a developmental assessment center. This lack of compelling legal pressure, however, does not negate the responsibility of the organization to make sure it is providing valid developmental feedback.

Another issue is the fact that many assessment center operators noted that they did, in fact, validate their center when it was first put into operation. The question arises, however, as to whether an organization should continue to execute validation studies on the ongoing process of assessment. One organizational spokesman commented that his company had been running an assessment center in the early 1970's, a center which they discontinued when there was a considerable downturn in their business. In attempting to revitalize that center in recent years, they have discovered that the dimensions, the job analysis from which the dimensions were developed, and even some of the exercises and assessment center data used are no longer appropriate for today's applicants or assessees. In other words, just as the economic situation for the company changed over the past nine years, so too had the nature of the job to which the assessment center had been targeted. If an assessment center is intended to select persons for a job, then an ongoing validation of that selection technique is reasonable from both a common sense point of view and an economic point of view (i.e., to insure continued return on investment).

One final validity issue which arises from these data concerns content validity. In the recent past, content validity was not a type of selection validation which was generally favored by the courts. The courts very definitely showed a preference for criterion-related validity as recommended in the EEOC guidelines of 1972. More recently, however, the uniform guidelines have suggested rather strongly that content validity, if adequately demonstrated, can be used as a persuasive and adequate defense of a selection technique. Many assessment center users seem to be relying very strongly on this implication. To the best of our knowledge, however, court tests of selection practices have yet to show that their acceptance of content validity has obviated the pragmatic organizational need to conduct criterion-related validity studies. This whole issue returns again to the question of what can, or will, be the economic return to the organization of using the assessment center method for selecting, promoting, or identifying developmental needs of its staff. Content validity can certainly give us information about the potential usefulness of a selection system, but it is an indicator of potential usefulness and not a measure of that usefulness.

Behavioral Dimensions Utilized

The conventional wisdom about the development of assessment centers suggests that assessment center designers should make a careful analysis of the job for which they wish to assess candidates. That job analysis should be done by using accepted and rigorous job analysis methods. We discovered that virtually all of the users in our sample did conduct some kind of job analysis. The specificity, rigor, and utility of those different job analysis methods varied considerably.

Some users rely on a literature review and a review of relevant research in order to "analyze" the target job for which they want to make assessments. While this, strictly speaking, could not be called a job analysis, it is a careful look at a job and an attempt to identify important categories of behavior which are needed for performance in that job. This "job analysis" methodology is most often used by assessment center designers who are developing a more generic kind of assessment center which will assess a number of people for a variety of jobs. For example, one university program which instituted the use of an assessment center for evaluation and prediction of job success among graduate students in psychology decided that the best way to identify the various fields and career opportunities for those students was to research the literature on careers in psychology. The program designers reviewed this literature, and then chose behavioral dimensions to be evaluated in the center on the basis of that research. One consulting organization has designed a relatively standard assessment center which it sells to organizations of all types (for use with management at all levels) based upon a similar kind of research and literature review intended to identify the generic broad-based skills needed for management. More importantly, educational applications of the assessment center method often rely on this kind of job analysis. An educational institution usually prepares students for a variety of jobs and for a variety of experiences in life. Therefore, in order to identify the critical skills to be assessed in the assessment center portion of an educational program, educators have tended to use available literature and research to identify broad-based skills rather than looking at specific jobs and their skill requirements.

Still other organizations have tended to rely on a consultant's files or experience for making a choice about the number and kinds of dimensions which can best be evaluated in their assessment center. In effect, little or no job analysis is done of the target job in the specific organization. One large supplier, for example, keeps on file descriptions of jobs and lists dimensions which were identified as important for them. It is unclear how many of these dimensions were directly generated by a job analysis or how many were derived (in an armchair manner) by looking at similar jobs and assuming that the apparent similarity in job descriptions indicated a similarity of dimensions for those jobs. A number of organizations lacking the resources for doing a more rigorous analysis have resorted to this kind of non-data-based choice of dimensions.

There are other user organizations and consulting firms which encourage and conduct more rigorous job analyses. These analyses involve a variety of structured, analytical methods for identifying and classifying important job tasks, behaviors, and/or outcomes. Such techniques include:

1. the use of a panel discussion with job incumbents or supervisors who have previously held the job which will be assessed in the assessment center;
2. the use of questionnaires listing potential job activities or skills, and having incumbents, supervisors, or other informed individuals indicate which of those activities or skills are required by the job and/or the relative importance of those skills for good performance;
3. the use of interviews with job incumbents, with supervisors of the target job, or with experts on the job (supervisors two or more levels above the job); or
4. the use of structured observations of actual incumbent performance on the job.

Each of these job analysis methods provides more detailed information about the target job than does either a literature review or use of information from similar jobs. Therefore, it is highly desirable to use the more rigorous methods of job analysis if one wishes to obtain detailed and specific definitions of the behavioral dimensions needed for the job in question, including any unique skills which are required for that particular job in that particular organization. Obviously, the use of a combination of methods is also very desirable, since each method represents a slightly different focus with respect to actual behavior on the job, and with respect to whether the technique measures characteristics of the individual or characteristics of the job.

Table 11 summarizes the information we gathered on this aspect of assessment center design. It is obvious that most organizations do some kind of job analysis, but the rigor of that job analysis is sometimes less than might be desirable. Our sample of users indicates that their job analysis techniques run the full gamut from armchair speculation regarding the important behavioral dimensions to rigorous combinations of multiple data sources. This is clearly an area of practice which requires considerably more standardization and research as to the best means for dimension creation.

The number of dimensions generated for assessment of a particular job varies considerably across centers and across organizations. For example, one large supplier, which has developed and distributes a generic center for use by management in a variety of companies, has generated 12 dimensions for its supervisory (or first level) management assessment program and 17 for its mid-level (or second level) management assessment program. This difference suggests that not all managerial jobs require the same knowledge, skills, and abilities. The number of dimensions generated for a particular job cannot be prescribed as such. However, it is interesting to note that most assessment centers evaluate somewhere between 6 and 20 dimensions with the mode being 11 dimensions (as represented in our sample). This suggests that, although there may not be a specific number of dimensions which can be prescribed for use in all assessment centers, there seems to be a practical range in terms of the number of dimensions which are identified and used by assessors in these centers.

One other interesting trend is apparent in Table 11. Educational applications of assessment centers tend to use a far greater number of dimensions. Those programs which use assessment centers for the evaluation of assessees who wish to be granted high school diplomas tend to use 60 or more dimensions. All 60 dimensions are evaluated by means of assessment methods. Other educational applications have a tendency to use 20 or so dimensions for evaluating performance or critical skills needed for jobs in educational settings. While there is not clear-cut evidence suggesting the reason for this phenomenon, it does appear that there is a tendency in educational settings to try to evaluate assessees in terms of their capabilities for a broader range of experiences. This brings us to the conclusion that as long as educational institutions try to educate individuals to be competent in the world, as opposed to being competent on a specific job, then the numerous and diverse dimensions which are evaluated in current educational applications of assessment centers are necessary and appropriate in terms of their breadth of coverage. However, this raises the very important issue of whether or not assessors can, in fact, identify, discriminate, and accurately evaluate 64 different behavioral dimensions. It has been suggested by the research literature that, in fact, it is very difficult for assessors to keep a large number of dimensions separate in their own minds (see Sackett & Hakel, 1979). Thus, the choice of these dimensions becomes a critical issue when one considers the implications of using the assessment center technology in educational settings. We will return to this issue in the research section of this report.

While the majority of assessment programs use somewhere between 6 and 20 behavioral dimensions for the organization and evaluation of assessee behavior, the labels of these dimensions vary considerably across jobs and across organizations. In our sample of organizations which were willing to share information about their assessment centers, approximately 35 were willing to share the names of the dimensions that they used in evaluating performance in those centers. From that group of approximately 35 organizations, we generated a list of 104 different dimension names (see Table 12). It is readily apparent that assessment center designers are attempting to measure a wide variety of human characteristics via the assessment center process. Several issues are raised by this abundance of dimensions.

First, this spectrum of dimensions varies considerably with respect to the behavioral specificity of the dimensions. Despite the historical debates regarding the use of underlying personality constructs, many of these dimensions clearly require that the assessor infer trait judgments based on some observable behaviors (e.g., genuineness, integrity, or flexibility). The dimensions also represent a mixture of attitudes, values, skills, knowledge, motives, and abilities. Thus, because of their variability in specificity, the key to the utility of a dimension lies in the behavioral indicators attached to the dimension label.

Second, while we could not feasibly elicit behavioral indicators within the content of our interviews with users, it is not unreasonable to suggest that one would find a considerable redundancy in these indicators across different dimension labels. For example, are the behavioral indicators truly different for "works for teamwork" than for "interpersonal sensitivity skills"? What are the differences in behavior required to rate "leadership" versus "management control"? Conversely, do different centers using the dimensions of "creativity" or "flexibility" really reflect the same behaviors?

Third, Table 12 does suggest that certain dimensions are much more commonly used in our sample of users. Twenty or more centers evaluate abilities to analyze and solve problems, planning and organizing skills, interpersonal sensitivity, resistance to stress, and oral communication skillls. Fifteen or more centers evaluate initiative, decision making, leadership, written communication skills, management control, and decisiveness. And ten or more centers evaluate persistence and flexibility. We would speculate that a large portion of this commonality is due to the impact of using common consulting organizations. Thus, the state of the practice is clearly influenced by these organizations towards the selection of a few basic dimensions reflecting the dominant functions of a manager's role.

In short, the state of practice with respect to the design and selection of behavioral dimensions for measurement is both complex and of widely varying quality. Few common standards can be identified, and few criteria exist for evaluating the best method of defining and describing the concept of a behavioral dimension. More work in this area is clearly justified.

Design of Performance Situations

Once one has identified relevant behavioral dimensions for a particular job, the next step in the design of an assessment center is to develop performance situations to elicit assessee behavior along the dimensions. There are a number of assessment techniques which appear in many of the centers in our sample. Table 13 lists the assortment of techniques that were reported in use by our sample of assessment center users. (A brief description of each of the more common exercise formats can be found in the appendix.)

As some behavioral dimensions are more common than others, so too are there assessment techniques which are more popular in practice. By far the most common performance situations are leaderless group discussions (based either on cooperation or competition of assessees with respect to the goal of the discussion) and the in-basket. Somewhat less common are various types of role plays (in which the assessee must assume a role to interact with others), structured interviews with an assessor, business games, oral or written presentations to the assessors, and case study evaluations. The use of psychological tests is notably much less common in practice. The reader can find a variety of other techniques displayed in this table, though these are much less commonly used in practice (e.g., peer ratings or work diaries).

The popularity of many of the more common assessment techniques can easily be traced to the origins of the practice in both the OSS program and the early research and practice at AT&T. However, Table 13 (specifically, the two columns on the right) would indicate that consulting organizations have exerted a great impact on common practices in this area. Many of these techniques have been packaged into standard formats and materials by consulting groups. In quite a few cases, users have employed these standard packages rather than designing the technique for the specific application or target job of the user's assessment center. These standard packages are often based on a generic group of management jobs, and thus are more applicable to centers designed more for job families (e.g., a particular level in the management hierarchy) than for one specific job. As a consequence, we did note that many organizations which have purchased such standard packages have begun to make in-house revisions in order to make the content of the situation more relevant to the specific job content within the organization. We suspect that this trend will continue.

To summarize, the current state of practice with respect to the design of performance situations would suggest that the typical assessment center relies heavily on four techniques: (1) the leaderless group discussion; (2) the in-basket; (3) role plays; and (4) structured interviews. Other techniques have been designed and used (e.g., business games, presentations, or case study evaluations), and practitioners have begun to display some ingenuity in the design of new techniques. However, greater attention must be paid to the evolution of criteria for the optimal design of individual techniques, and for the optimal mixture of various types of techniques in any particular center. Such criteria would clearly contribute to advancing the state of the art, and would combat the apparent tendency to employ a technique simply because of its availability from materials suppliers or because of its common usage in other centers.

The Composition of Assessor Teams

While the exercises chosen for use in an assessment center provide the opportunity to observe job-relevant behavior, it is the assessor team observations and judgments which ultimately determine the validity and utility of the assessment center as a method of gathering information on potential job performance.

Typically, assessment centers use supervisors from the organization as assessors. These supervisors are typically drawn from one to three management levels above the assessee's level in the organization. As demonstrated in Table 14, however, there are a very few assessment centers which use peers of the assessee as assessors of performance in addition to supervisors. Another alternative is the use of psychologists as assessors. Large organizations with personnel research staffs, or psychologists responsible for the primary design and administration of an assessment center, sometimes use these groups as assessors--though this is far less common. Lastly, there are a few centers which use some "outsiders" as assessors. The term "outsiders" is used here to indicate that the individual who sits as an assessor is not a direct supervisor of the participant, nor a peer, nor a staff psychologist with the organization. Supervisors from other parts of the organization (other departments, staff functions, etc.) as well as professors, members of the community, and business people with knowledge of the target job, are sometimes called upon to act as assessors in assessment centers. As Table 14 would indicate, the current state of practice is quite varied with respect to the composition of assessor teams from these various sources. Several issues are germane to the selection of assessors from these sources.

As already noted, peers are seldom used as assessors. This may be a function of a number of issues. First of all, peers may not have the knowledge of the target job which would facilitate their being able to judge and evaluate each other's performance in the assessment exercises. Second, there might be some competition between a peer assessor and his or her associate with respect to the target job. And third, peers' judgments are definitely an uncommon basis for selection and promotion of people within an organization. Therefore, the use of peers tends to appear either in situations where there is no hierarchy that would permit a supervisor to have more information about the job than a peer; or in situations where assessment is done for developmental purposes and therefore would not engender competition between peers. For example, our sample includes one university which employs the assessment center process for the selection of faculty. It is traditional for peers to select faculty members, rather than deans or administrators who usually have less specific knowledge of the requirements of the faculty member's job. Two other applications using peer assessors are of a developmental nature. One is conducted by a large corporation which uses its assessment center process strictly for developmental purposes. In this particular organization, assessees give each other feedback on their performances in assessment exercises. Nonetheless, the more traditional pattern holds when assessment reports are written up detailing the individual's performance; these reports are written by the assessor staff only. The second developmental application is an assessment center run for evaluating graduate students in a business school. The peer assessors are two members of a six-person assessment team which cooperatively evaluates student performance and writes the reports.

Typically, assessment centers use supervisors as assessors of performance in the assessment center exercises. Usually the supervisors are two or more levels above the job incumbent. This difference of two or more levels reflects a concern in the organization that a worker's immediate supervisor not be the one who evaluates him or her. This is done so that prior knowledge of the candidate does not contaminate judgments made in the center. Note, however, in Table 14 that the number of supervisory assessors from different levels is unclear. This is an artifact of the way in which information was shared with us. Practitioners generally tended to indicate that supervisors served as assessors without specifying at what level those supervisors functioned in the organization. For those cases where level of supervision was not indicated in the raw data, "level unspecified" is recorded in Table 14. Briefly, the rationale for using supervisors as assessors is that they have some first-hand knowledge of the job which is the target of the assessment process and they are individuals whose judgment is valued by the organization. Using supervisors who have organizational credibility lends credibility to the assessment process.

Last, and perhaps least, there are organizations who use staff psychologists as assessors. In developmental centers, the use of staff psychologists as assessors may have considerable face validity and acceptance among participants. However, in assessment centers where judgments are made of management skills, and promotion decisions are the product, psychologists are less likely to have credibility with assessees. Past research has suggested (see Taft, 1959) that psychologists are no better at perceiving and evaluating behavior in the assessment center than are managers. Therefore, most organizations try to use managers instead of psychologists as the staff of their assessment center. (Also, most organizations do not employ a sufficient number of psychologists who could staff their assessment centers.) This factor has contributed substantially, we would speculate, to the growth in applications.

Assessor Training

The quality of an assessor's observations is fundamental to the effectiveness of the assessment center process, as these observations are the foundation for the rating of behavioral dimensions and for the overall competency judgments derived at the conclusion of the center. Consequently, one could reasonably argue that the effectiveness of the technology hinges greatly on the quality of assessor training.

Table 15 displays common practices in this area according to our sample of users. Several salient facts emerge. First, the length of these training programs varies from four hours to an entire week, though the average length would appear to be about three days. Only a very few organizations did not offer any training, though many organizations did not, or could not, clearly specify the nature of their assessor training programs. Many current assessment program administrators with whom we talked were not involved in the initial phases of training design and appeared to lack detailed information regarding the training. For whatever reason, there did appear to be some insensitivity to the critical importance of assessor training design by some users.

Second, the goals of the assessor training process are fairly common, namely: (1) to familiarize assessors with the behavioral dimensions and specific performance situations to be utilized; (2) to practice observing behavior in the situations and to practice using the dimensions to rate the quality of the behavior observed; and (3) to experience the dynamics of assessor team discussions for reaching overall judgments about an assessee. The procedures typically used to accomplish these goals are also fairly standard in practice, and include such activities as role-playing the exercises, actually participating in the exercises, reading supporting materials, or practice observation and rating of performance in the situations (including the performance of other assessors or "sample" assessees--either "live" or on videotapes). We would furtner speculate that standard practice in this area has been highly influenced by the relatively standardized assessee training packages offered by the major assessment center consulting firms. These packages do provide an organization with a fairly comprehensive and well-structured set of activities for training assessors in the observation of behavior in the performance situations.

Third, and perhaps most troublesome, is the observation from Table 15 that few organizations in our sample require assessor certification of competence, or offer refresher training for assessors who have not participated in a center for a lengthy period of time. The issue of assessor certification has become a "hot" item among practitioners, and truly represents a difficult dilemma. On the one hand, the assumption is often made that a skillful manager already possesses many of the basic skills required to be a skillful assessor, and training in observational techniques simply accentuates and refines these skills for the specific purpose of the assessment center. On the other hand, the argument for certification rests on the assumption that not all individuals are equally good at organizing and evaluating the behavior of others; therefore, only the best should be used in the assessment center. The problem arises in the mixture of the two assumptions; that is, if one does not certify a potential assessor as being "good enough" to serve on an assessor team, what does this mean with respect to his or her managerial skills on the job?

Most organizations would probably wish to avoid passing judgments on their assessor staffs, given the need for line managers to participate as assessors in the program, and given the potential organizational difficulties which might arise in rejecting assessor candidates, not to mention the cost of training an unsuccessful assessor candidate. Furthermore, despite the packaged training programs available, little research or theory has been developed which (a) succinctly pinpoints the key skills and abilities required for the assessor role; (b) concretely specifies the form which the training activities ought to take in order to build these skills; or (c) provides measurable criteria upon which to base a certification or rejection decision. As a consequence, it would appear that very few organizations certify, or reject, assessors after the training program. (This point is further supported by the results of an informal poll of users conducted by one speaker at a recent conference on the assessment center method.) The current state of practice would be considerably strengthened were more attention paid to this issue.

A fourth, and final, issue emerges from our interviews with users regarding assessor training (though this point does not appear in Table 15). Many users report having used one of the major consulting organizations to design and/or conduct the initial assessor training program, but few report any support mechanism within their organizations to facilitate the continuation of this more rigorous kind of assessor training. It would appear that formal assessor training programs often dwindle somewhat, once the consulting organization is no longer active in the administration of the center. For example, several users reported that their assessors were originally trained by one of the consulting organizations; however, as individuals dropped out of the assessor ranks over a period of time, their replacements often received their training by observing in a quasi-assessor role, by reading materials or exercises, by observing an ongoing center, or by receiving a more limited version of the elements of the original training program. While such experiences may well be as effective as the more elaborate and expensive standard training programs offered by consulting firms, there is little evidence available for substantiating this effectiveness. In short, quality control standards for asessor training would appear to be greatly lacking in many organizations. This should be a key target of future practice.

In summary, the state of practice with respect to assessor training is of varied quality. Nearly all organizations provide some kind of assessor training which typically involves some role-playing of the assessor's role prior to the actual center. However, few organizations have closely examined questions concerning the basis for the effectiveness of assessor training programs, or have designed clear quality control procedures for ensuring assessor competence. The need for specific training standards based upon careful research is readily apparent; in their absence, the effectiveness of the assessment center process can be seriously compromised.

Methods for Selecting Assessees

Another very important issue for consideration in the discussion of the assessment center process is the means by which assessees are selected to be participants in an assessment center. The actual selection of assessees varies across organizations, as reflected in the data presented in Table 16. The more traditional means of selecting assessees has been for the supervisor within the organization to nominate or recommend that certain of his or her subordinates be included as assessees in an upcoming assessment center. This traditional method of selecting assessees has been modified somewhat in recent years to include more opportunities for self-nominations.

As described previously, selection and/or promotion have often been the primary purpose of assessment center applications in industry. Under those circumstances, it only made sense that supervisors would nominate those people whom they thought had the greatest potential for moving up into management (or some other new job). The method of assessee selection follows the traditional hierarchical arrangement of most organizations. However, the use of assessment centers has greatly diversified over the last 10 years. Assessment centers are being used more than ever for developmental purposes, and individuals within the organization have been given increased opportunities to recommend themselves for this developmental process. This is analogous to allowing an individual to sign up for a training program. Self-motivated workers certainly are in a good position to recommend their own candidacy for development, and are likely to follow through on developmental plans emerging from the process.

In addition, in the last few years organizations have tended to avoid relying solely on supervisory nominations as a basis for selection or promotion decisions. It can be difficult to demonstrate that supervisory nominations are not biased against legally-protected classes. In fact, the courts have ruled that management nominations cannot be assumed to be valid predictors of future job performance unless they can be shown to have no adverse impact on minority group selection. If an organization requires participation in an assessment center as at least part of the employment decision process, then nomination to the center as a prerequisite to the assessment center process requires evaluation. Consequently, many organizations have turned to the use of a self-nomination procedure to offer equal opportunity to all applicants.

Nevertheless, relying on self-nominations can lead to a deluge of applicants for the center, many of whom may possess few qualifications for the target job. Some organizations have opted to allow full participation; others, in an attempt to control costs and minimize the time needed for decision making, have tried to design some initial screening mechanism to narrow down the applicant pool. This mechanism might consist of some combination of self-nominations, supervisory recommendations, personnel office recommendations, or qualifying interviews and tests. We know of at least one organization which has gone so far as to develop a paper-and-pencil test battery to be administered to assessment center applicants in order to predict the probability that the applicant would receive a high overall rating from the assessment center process. The results presented to date show a fairly high correlation; however, one must ask whether such a paper-and-pencil screening device, if highly correlated with assessment center performance, does not undermine the need for the assessment center itself! The use of screening procedures for assessee selection is a relatively new development in practice and will surely continue to evolve as organizations attempt to minimize the costs associated with full-scale assessment center implementation for selection purposes when facing a large applicant pool.

Table 16 would also suggest that the selection of assessees in educational applications typically follows the self-nomination format, though several universities either strongly recommend or require that all students participate in the assessment program as part of their educational commitment. The format would appear to vary based upon the degree to which participation in the assessment program is a formal part of the student's performance evaluation, or merely a developmental adjunct. Clearly, practice in this area has not been standardized.

Assessor Observations, Ratings, and Competency Judgments

This aspect of assessment center technology, as currently in practice, is difficult to represent in tabular form due to the multitude of variations in practice. What we can present is a fairly brief verbal description of the common themes in practice.

The ratio of the number of assessors per assessee varies considerably across our sample of users--ranging from a ratio of one-to-one to a ratio of one-to-six--though the more common ratio would appear to be one-to-two. (This ratio also seems to be typically encouraged by consulting organizations). This often means that the assessor team is composed of six to nine individuals and the assessee group of twelve to eighteen participants. The lower end of these ranges appears to be more typically preferred due to the complexity of the observation and rating procedures necessary in the process.

Within specific performance situations, one or more assessors may observe each assessee, though some formats (e.g., leaderless group discussions) call for one assessor to observe more than one assessee. Across the various performance situations composing a center, assessors are usually assigned to different assessees. Therefore, by the conclusion of the center, each assessee's behavior has usually been observed by as many members of the assessor team as possible.

In observing assessees in the performance situations, assessors are usually provided some form for the recording of behaviors exhibited. In many instances this form simply demands a written description by the assessor of any action displayed by the assessee--ranging from nonverbal cues to specific verbal comments or physical activities. In a few instances this form may demand some initial scaling of the behaviors observed on a scale anchored in terms of specific behaviors (this is called a behaviorally-anchored rating scale). Assessors may, or may not, rate behavioral dimensions following each performance situation, though the common wisdom from consultants is that such ratings should be withheld until the final assessor team meeting.

The final assessor team discussion lies at the heart of the assessment center technology. It is during this meeting that the various observations and dimension ratings by assessor team members must be discussed and reconciled. The focus of this discussion may vary somewhat depending upon the data available. In the case where the behavioral dimensions or some subset of dimensions have been rated within each performance situation for each assessee, any discrepancies in these ratings among the different assessors who observed the assessee in the different situations must be reconciled. The rating scales are typically composed of 5-point to 9-point scales. Any discrepancy in assessor ratings which exceeds 1 or 2 points usually should be reconciled through discussion until consensus is reached. Obviously, some of these discussions can become heated, even though emphasis is usually laid on grounding the ratings in specific observable behaviors.

In the case where the behavioral dimensions are not rated for each assessee until this final meeting, the discussion usually centers on the behavior reports written by each assessor who observed the assessee in the various performance situations. Following the reading of the reports, each assessor rates the dimensions for that assessee. At that point, further discussion is required if discrepancies are identified. Once again, consensus is usually demanded.

Once the behavioral dimensions are rated for each assessee, the assessor team is usually required to reach an overall judgment of competence upon which some organizational action may be taken, though this may vary with the purpose of the center. These overall judgments may be attained by averaging the various dimension ratings, by requiring the assessors to reach consensus on an overall rating, or by some subjective or objective weighting of the various dimension ratings which reflects the more important skills required in the job. This final overall judgment is typically placed on a 2-point to 5-point scale indicating degree of potential, or of acceptability for selection or promotion (once again the nature of this scale depends upon the purpose of the center). In centers with a more developmental focus, such overall ratings are less common; rather, the assessors are required to take special effort to document the specific behaviors which may be indicative of strengths and weaknesses in performance in order for the assessee to build a training or development plan. This latter procedure is much more common in educational applications.

From our discussions with our sample of users, and our own experiences with this aspect of the assessment center technology, several important issues can be raised concerning current practice. First, since the assessor team discussions usually center on the categorization of behavior to dimensions in order to rate behavior, the degree of specificity inherent in the definitions of the dimensions becomes of critical importance. Assessor team discussions can often digress when the team does not share a common point of reference for ascribing behaviors to dimensions. For example, several centers use the dimensions of "decision making" and "decisiveness," with the former referring more to the ability to form judgments in a sound fashion and the latter referring more to the relative speed and authoritativeness of a decision. These dimension definitions often appear to be sources of confusion for assessors, and can lead to unnecessary debates within the assessor team at the time of rating.

The more important aspect of this issue is insuring that a specific assessor team, or a number of teams in the same program, utilize the dimension definitions to identify and rate the same set of behaviors originally included under the dimension by the designers of the center. A common difficulty often cited in the assessment center literature is that the assessors do not adhere to the behavioral definitions of the dimensions, but rather import their a priori conceptions of "what it takes to be a good manager," and use these concepts to rate the dimensions or to make overall competency judgments. This problem once again emphasizes the fundamental importance of assessor training as a vehicle for eliminating prior stereotypes and for grounding judgments in specific behaviors.

Second, a variant of the first issue lies in the problem of assessor judgments being based on relative or absolute performance. Common wisdom, and many training programs, emphasize that an assessee's performance should not be rated with respect to the performance of the other members of the assessee's group; rather, the rating scales are intended to reflect an absolute level of performance (i.e., a 5-point performance in one group would be a 5-point performance in another group.) In practice, however, this can be an extemely difficult point to establish with assessor teams. It requires heavy emphasis in the assessor training, with concrete examples provided for each level of performance, and it requires close monitoring and controlling by the center administrator. There is at least some research which suggests that this problem can be controlled, but far greater research is required before definitive conclusions and standards can be reached (see Section IV for further discussion of this point).

Third, much more research would be useful to establish other standards in this area of practice. For example, what is the optimal ratio of assessors to assessees, and does the composition of the assessor team require special design consideration in relationship to the composition of the assessee group (i.e., on the dimensions of age, sex, race, etc.)? How many assessors should observe an assessee's behavior in specific performance situations in order to obtain sufficient reliability and validity? Are any differences in ratings apparent when assessors rate dimensions in each performance situation versus when they rate all dimensions at the end of the center? What is the best method for summarizing multiple observations and assessments--statistical procedures or assessor judgment? (See Section IV for the inconclusive evidence on this point.) Each of these questions is highly deserving of research; advancing the state of practice will clearly hinge on the answers.

Fourth, and finally, it is obvious that the key to the effectiveness of the assessment center method lies in the quality of the assessor team discussions concerning their observations, ratings, and judgments; yet this aspect of the technology has received very little attention in research. More attention must be paid to the determinants of the quality of this group discussion, and the factors associated with the group dynamics, per se, which accentuate accuracy in the group judgments. There is a voluminous literature on small-group behavior which could be brought to bear upon this problem and help guide future research. The outcome of such research would be extremely useful to center administrators in establishing guidelines for controlling the dynamics of these group discussions, and would further increase the accuracy of the judgments made in these situations.

The Feedback of Assessment Center Results

Most professional ethics guidelines, including those created for the assessment center field, establish the rights of assessed individuals to some form of feedback on the assessment results. As indicated by Table 17, sharing of feedback is fairly common practice among our sample of users. However, there are some variations in practice in the manner and form with which this process occurs, and in the extent to which other members of the organization are allowed access to the results.

Most often the assessment center data is provided to the assessee in some form. In many centers, each assessor is required to generate a written report on each of a subset of assessees which reflects the behavioral examples underlying the dimension ratings and the rationale for the overall judgment. In some cases this responsibility might lie with the on-going staff or administrators of the center. The actual sharing of this report with the assessee also varies. In some organizations, the assessee receives the report, or a summary, in the mail, though this practice is not widely advocated. In others, the assessor, or center administrator, meets with the assessee, following the conclusion of the center, to discuss the results orally, with receipt of a written report being optional. In still other organizations, the assessee may be called in later to meet with an assessor, a supervisor and/or a staff counselor to discuss the report and, once again, may or may not receive a written copy of the report. In many of these organizations, the assessor may reserve the option not to distribute feedback, in which case the assessee would receive the report only upon request.

When the purpose of the center is for selection and promotion, the results are most commonly provided to the assessee's supervisor or to the individual responsible for the selection decision. Practitioners often emphasize that the assessment center results should serve as only one input to the decision-making process, and should be considered in relation to past performance on the job. It is not clear how closely this admonition is followed in practice, nor is much information available regarding how these supervisors from outside the assessment process use the information generated in the assessment center. This would appear to be a fertile area for future research with respect to the organizational consequences of assessment center implementation.

We did find some organizations which provided feedback to the assessee's supervisor even when the purpose of the center was for training and development. While this often reflected a concern for sharing information with the supervisor about his or her subordinates, it can present something of a dilemma for both the assessee and the supervisor. Both may actually benefit from the discussions of personal strengths and weaknesses and the insights generated for future developmental needs. However, the possibility remains that future selection decisions might be affected by data-sharing which is reputed to be developmental only. Consequently, it would seem to be critical, in developmental assessment programs, for the assessee to have ultimate control over the dissemination of the results of his or her performance in the center. This is clearly not always true in current practice.

The least common method of assessment center feedback is to provide the assessee's results to a counselor or staff person in charge of training and development. Despite the fact that many organizations are beginning to use the assessment center process as a tool for individual career development, few have provided the internal resources to help the assessee to use the results actively. The assumption commonly made is that the individual assessee will take the initiative to follow up on developing the weaknesses identified in the center. As a consequence, the organization runs the risk of investing considerable expense in the design and operation of a center without facilitating a return on its investment through concrete goal setting and action planning with the assessees in terms of future development.

Table 18 further documents this point about follow-up from the assessment center process. For our sample, the data indicates that individuals are three times as likely to receive feedback on their results than they are to receive any guidance in developing a plan for improving their career possibilities. Only about one-fourth of those assessees who receive feedback after their participation in the assessment center go on to work out a developmental plan. (While there may be individuals who go on to develop a plan on their own initiative, Table 18 specifically identifies those programs in which a developmental plan is required.) More common than the individual developmental plan requirement is the sharing of information and working on developmental plans with one's supervisor. This is also a more traditional kind of sharing following the normative hierarchical structure of most organizations. The least common strategy for assessment follow-up is the provision of feedback to the individual as well as staff resources for the on-going monitoring of a developmental plan for that individual following his or her participation in the center.

We have examined the state of practice in the feedback of assessment center results primarily from the point of view of the individual assessee; however, we must also comment on the use of these results by the organization. Table 19 summarizes this information for our sample of users. As might be expected from our previous comments, nearly all of the organizations report some form of feedback to the assessee, and quite a number report use of the data for various personnel decisions. About one-half of the sample report specific procedures for the design of developmental plans, but only one-third report any organizational retention of the results for purposes of research or validation efforts.

It would appear that many organizations have struggled with the issue of retaining assessee results in an organizational data bank. With increasing concern over the invasion of privacy when someone gains access to such computer-based data banks, many organizations seem to have opted for not retaining any data on a specific assessee. In contrast, in those organizations (such as AT&T) which have developed massive data banks, considerable precautions have been required in order to insure the confidentiality of the information beyond its initial use for a specific selection decision, or a development plan generation. We identified at least two organizations where the issue of data retention by the organization created serious resistance to the process within the organization. In one case, rumors were widely circulated within the organization that assessee results from a developmental center were placed in a data bank being tapped by executives from all over the organization in order to form candidate pools for different jobs. This may well have violated the prior consent of the assessees regarding organizational access to the data. Similar problems may result when assessment data are fed back to an assessee's immediate supervisor, who then shares the results with other managers for other decisions which may, or may not, be appropriate applications of the information.

Practitioners are highly sensitive to these issues, and it is likely that these problems represent isolated cases where users are minimally sophisticated. Nevertheless, the important point remains that the organizational use of assessment center results can be beset with difficulties. It would appear from our discussions with users that greater attention must be paid to this aspect of the technology in the future, particularly at the early stages of center design. Without pre-planning, the data-base required for future research may be seriously constrained, and the centers themselves may encounter legitimate, albeit unnecessary, resistance from members of the organization who are sensitive about how assessment data will be used and by whom.

Further Consequences of Assessment Centers for Participants

The nonempirical assessment center literature is replete with references by advocates to the unintended positive consequences of implementing the assessment center concept. Skeptics have offered an equal number of hypothesized negative consequences. Unfortunately, little data is available to support the conclusions on either side of this debate. Nevertheless, we believe it is important to, at least briefly, examine some of the more commonly cited consequences in order to fully convey the state of the art to the reader.

One of the most commonly cited positive consequences of the assessment center process is the potential benefit which can be derived by supervisors from participation in assessor training, and from acting as observers and evaluators in the process of assessment. The nonempirical literature suggests that assessors develop greater observational skills in the assessment center which they then transfer to other management functions outside of the center. To the best of our knowledge, however, there is little or no empirical evidence demonstrating that assessors do, in fact, (1) improve their observational skills over some basic talent for observation which they bring to the assessment process; and (2) transfer any skill gained or used in the assessment process to other management functions outside of that setting. The evidence is anecdotal at best. Organizations consistently report that their assessors do gain some benefits incidental to participation in the assessment center process. This seems entirely reasonable, but we would prefer documentation of this point.

For example, one organization from our user sample is specifically committed to using the assessment center as a developmental experience for participants. That firm allows upper level management to participate as assessees in the process, despite the fact that the assessment center is designed to select people into the job that they have already attained. This particular organization reports that a number of upper level supervisors find the experience to be developmentally beneficial. Therefore, upper level management people may opt for participation in the assessment process strictly for developmental purposes. Some developmental benefits are presumably gained by participation in the process, beyond the specific feedback around job-related skills. Benefits are not documented beyond subjective testimonials.

A second positive, albeit biasing, consequence of assessment centers often cited is something called the "crown prince" effect. The assessment center literature has suggested that an individual who participates in an assessment center, and receives positive ratings, may later be treated as a "crown prince" on the job. More specifically, the literature suggests that when a positive rating from an assessment center is communicated to a supervisor, that supervisor may treat the highly-rated participant with extra deference. That participant may, in turn, receive extra challenges, extra rewards, and earlier-than-typical promotion. This is essentially a type of self-fulfilling prophecy.

There has been little evidence that this crown prince effect really does occur. One of the organizations in our sample, however, has recently demonstrated a crown prince effect following positive assessment in their assessment centers. Specifically, they examined change in salary growth, merit increases, and annual job changes for former assessees. After a period of two years, assessees who were rated high and those who were rated low did not differ on either the merit increases they had received or the annual number of job changes, but did differ in terms of salary growth. More specifically, those individuals who were rated high by the assessment center were more apt to have received large salary increases than were those assessees who had not been so highly evaluated. The assessees rated high and low had been equal in salary prior to their assessment center experience, but were significantly different in the amount of salary they received approximately two years after assessment. Thus, at least in this one organization, there is some evidence of a crown prince effect. It is a positive consequence of assessment in the sense that this may bring high rewards to individuals with great potential to contribute to the organization. It may, however, be seen as a negative consequence in the sense that those who receive low ratings in the assessment center were not differentiated from highly rated participants in terms of job changes and merit increases received. This may reflect the basis upon which merit increases and job changes are awarded, or it may reflect the true nature and significance of differences between the two groups in performance. That is, either the organization is awarding merit increases to mediocre performers, or there is no real difference between those rated high and those rated low by the assessment center, when it comes to their actual job performance.

A third suspected positive consequence of the use of assessment centers is the realistic preview that the assessment center process gives a job candidate. If assessment centers do, in fact, provide situational exercises which reflect the requirements of the job and have some face validity in terms of the actual content of the job tasks, then those tasks in the assessment center may be seen as realistic previews of the job. Several organizations have reported that the assessment center's realistic preview has positive consequences for the participant and the organization. A couple of different insurance firms, a large technical corporation, and three different universities all report using assessment centers for the purpose of development or selection and also as a realistic preview of the job. One organization which had a chronic turnover problem specifically chose the assessment center because it was a realistic job preview. They expected to reduce the turnover among new job hires by giving them a balanced set of expectations about the job. Another organization showed a short film of the actual day-to-day frustrations in the target job, at the very outset of the assessment process, in order to give a realistic preview. The three educational assessment centers which claim that realistic job expectations benefit assessment center participants all undertook their assessment centers for the purposes of both (1) evaluating and developing skills which would be needed on the job, outside of the academic setting; and (2) giving the students some realistic expectation about the job for which they had chosen to train. The notion among these educators was that academic evaluation of performance in course work was not necessarily highly related to performance on the job. They felt that the situational exercises of the assessment center were a more realistic preview for the student, and would help the student to determine whether or not he or she would be happy in or capable of applying academic knowledge to the actual situations he or she would confront on the job.

Another commonly cited positive consequence of the assessment center process is the greater degree of legitimacy ascribed to the selection decisions by assessees or other managers. There is a reasonable amount of empirical evidence (e.g., Kraut, 1972; Boche, 1977; Dodd, 1977; Nirtaut, 1978) which suggests that most participants in an assessment center accept the process and view it as more face valid than other techniques, though Nirtaut does note that about 14 percent of the organizations he surveyed reported some adverse reactions to the process. Future research might well advance the state of the art if these few adverse reactions were more closely studied.

The most commonly suggested negative consequence of participation in an assessment center is something which is called "the kiss of death." This phrase is used in the nonempirical literature on assessment centers to suggest that an individual who attends an assessment center, and receives a poor rating on his or her potential or job-related skills, will essentially be getting the "kiss of death" from the organization. Assessment reports, once fed back to supervisory and management people, are hypothesized to lead to negative consequences for the individual participant in that supervisors will fail to promote, fail to give opportunities to, and fail to recognize the achievements of individuals who have received negative ratings in the assessment center. If a low assessment center evaluation of the individual is indeed accurate, then it is likely that there will be little cause to praise the assessee back on the job. If, however, the assessment center evaluation is in error (for whatever reason), then negative consequences deriving from the negative assessment report may be extremely unfair. While this all seems logical, to the best of our knowledge there is little empirical evidence that poor performers who receive negative assessment ratings are treated any worse by managers who see those ratings than they would be if the supervisors reacted to them on the basis of their job performance alone. There is, at least, some evidence (Kraut and Scott, 1972; Nirtaut, 1978) available to suggest that the "kiss-of-death" phenomenon is not a particularly powerful effect, although some turnover and attrition have been experienced by a few organizations as a consequence of assessment center results for specific assessees. Certainly more research would be valuable in this area in order to assess the degree to which the kiss-of-death truly operates and/or is justified in terms of actual job performance by assessees.

Another commonly cited negative consequence is that the assessment center process creates undue stress for the assessee, stress which can seriously alter his or her "natural" job-related behavior, or which may have long-term emotional consequences for the assessee. There is simply no evidence available to support or deny these contentions. Certainly the assessment center does raise anxiety in assessees, particulary when the center is designed for selection purposes. It is not clear that this anxiety has any short-term or long-term effects, and practitioners can provide a compelling rationale that such stress is a part of the target job, so its inclusion in the assessment center is appropriate. This would be an intriguing area for future research on the effects of anxiety level upon performance.

In summary, the unintended positive and negative consequences of the assessment center process are widely discussed in the current literature, though little empirical evidence is presently available to substantiate many of the claims. Future practice would undoubtedly benefit from a closer examination of these issues.

Consequences of Assessment Centers for the Organization

While there is at least some discussion, and some research evidence available regarding the effects of assessment centers on participants, there is very little information available regarding the broader effects of assessment center use upon the organization per se. It would appear that the assessment center concept does hold the potential for affecting the entire human resource development function within an organization. (In fact, this impact was a major theme of the 1979 International Congress on the Assessment Center Method.) Since this extension is likely to represent the next stage in the evolution of the concept, we feel it is important to offer some speculations regarding possible organizational consequences of assessment center implementation.

The utilization of an assessment center for the purposes of selection or promotion quite obviously represents a significant change in the means by which organizations reach one of their most fundamental decisions--the placement of personnel. A common fear often expressed at the early stages of assessment center design is that the use of assessment results reduces the supervisor's traditional prerogative to select or promote whom she or he wants. We suspect that this fear represents only a partial truth. Many organizations establish, as a matter of policy, that the results of the assessment center should be but one input to the supervisor's decision making, and the supervisor can overrule the assessment results with other data available on actual job performance. Nevertheless, supervisors may find it extremely difficult to ignore the results of the assessment center, particularly if a selection or promotion decision counter to the assessment findings leads to a job failure.

Furthermore, the supervisor may experience some peer or boss pressure either to use or ignore the assessment results. Consequently, the implementation of a center for employment decisions must pay heed to the prior organizational traditions regarding such decisions and include plans for facilitating what amounts to a cultural change.

At the heart of this issue is the fact that, in many organizations, the criteria and procedures for employment decision making are relatively mysterious, nonpublic, and subject to individual supervisory bias (i.e., two supervisors may not emphasize the same criteria or use the same procedure for making an employment decision). In contrast, the assessment center provides an explicit procedure, a relatively public set of criteria based on observable behaviors, and multiple judges to avoid individual bias. Thus, the fundamental basis for the perceived legitimacy of the organization's decision making is altered with the implementation of the center as part of the process.

This shift in the basis of legitimizing employment decisions also represents somewhat of a challenge to the typical, and traditional, hierarchies of most organizational bureaucracies. The assessment center concept does de-emphasize the importance of using a single manager as supreme judge. It does create pressures to base such decisions on publicly verifiable data, regardless of the position of the manager in the hierarchy. In short, the emphasis is on the power of the data from the process, not on the power of the manager in the hierarchy.

In this sense, the assessment center concept represents a striving for equity in the context of hierarchy. Therein lies a source of tension and practical consideration for every user of the assessment center concept as a basis for employment decision making.

The evolution of the concept into selection practice has consistently displayed the impact of this issue. The trend in practice to compose assessor teams of supervisors rather than peers represents one simple example of its impact. The presentation of assessment results to the assessee's supervisor, and the sensitive issues created by this feedback, is another example. Based on our conversations with users and practitioners, the trend towards more developmental applications might also be viewed as partially a function of the organizational resistance associated with implementation of a center for selection purposes.

Employment decision making on the basis of an assessment center clearly has implications for other employee appraisal procedures in use by an organization. Many organizations have developed at least some minimal procedures for the periodic evaluation of employee performance by supervisors. Such procedures are often tied in some way to the planning or goal-setting systems (as in Management-by-Objectives) or to compensation and reward systems. Many practitioners have recognized, in conducting validation studies, that using assessment center results to predict these supervisory ratings is a hazardous

131

strategy. These performance appraisal ratings are often unreliable, are often based on quite different rating dimensions, and/or are often the result of varying levels of actual observation of behavior on the job. As a consequence, the use of an assessment center for selecting an employee into a job (or even as a developmental experience for the employee) can create inconsistencies with the evaluations produced on the job. Such inconsistencies can be particularly troublesome if employees come to expect that their goals and rewards for performance should be based on the same criteria or process evident in employment or development decisions from an assessment center.

Some organizations have experimented with the idea of including the behavioral dimensions from the assessment center within the standard performance appraisal documents provided to the supervisor by the organization. This seems to be a reasonable strategy, but can be fraught with difficulties if supervisors do not have the training to use the dimensions in rating observable behaviors, or the skill to provide behavior-based feedback to the employee. This is an area of future practice which will surely bring new innovations as organizations attempt to reconcile the rigor of assessment center procedures with the far less rigorous practices apparent in typical performance appraisal procedures.

The implementation of an assessment center for developmental purposes would appear, on the surface, to be much less problematic with respect to the consequences for the organization--certainly a common remark by many users and practitioners. Once again, this may be only a partial truth. Developmental applications are likely to be less troublesome than selection applications with respect to possible legal repercussions for the organization, and they are likely to have less direct impact on prior organizational traditions in decision making. Nevertheless, developmental applications may exert subtle influences on the organization to provide greater resources to the human resource development function. This impact may arise from two sources.

First, because developmental applications are of a more recent origin than selection applications, it is not yet fully clear in current practice whether the design of the center requires significant change from common selection-oriented practices. Boehm and Hoyle (1977) have suggested, for example, that assessment for development will require greater emphasis on the reliable measurement of the behavioral dimensions and on the validity of these dimensions for predicting actual job performance. This is because the feedback from the center must be able to validly discriminate an assessee's strengths and weaknesses in specific skill areas. Thus, developmental centers might require even better job analyses, more complex performance situations, or even more assessors--all of which will require more resources from the organization.

Second, the provision of developmental information to an assessee can activate further needs for training and development.

As reported earlier, some organizations are now beginning to provide staff counselors to help assessees construct a developmental plan based on the assessment results. These plans can lead to expectations that the organization ought to provide, sponsor, or pay for the further training needed to correct certain deficiencies. Furthermore, it is not entirely clear that all the characteristics measured in an assessment center are trainable. An organization will require better information on this issue in order to allocate its resources effectively to employee development; and trainers may be confronted with the problem of demonstrating the effectiveness of their programs to produce real development along the assessment dimensions. Thus, the organization may have to devote more resources to the design and implementation of training programs for employees with measurable impact. It is our belief that most organizations are a considerable distance from this goal.

There is one final means by which we can assess the consequences of assessment center implementation for an organization; namely, by examining cases in which an organization has discontinued its use of a center. Such cases are rarely, if ever, discussed in the literature. In our sample of users, such cases were indeed rare, though not nonexistent. Several rationales for discontinuing a center were apparent in these cases.

One organization reported dropping the use of an assessment center because it was not sufficiently job-related. The lack of job-relatedness for this particular center derived from the fact that it was clinical in nature, in the sense that job skills and traits were evaluated, but those traits were identified on the basis of "armchair" speculation of what personality traits were most important for the job. Candidates for the job were then assessed in a clinical manner, which concluded with judgments on candidate personality, motivation and traits, rather than on candidate performance of job-related behaviors. It should be noted that this organization is now in the process of conducting the job analysis and preliminary steps required to build a more rigorously job-related assessment center. The company intends to revitalize the assessment center and, hopefully, to provide a more valid system than was previously in place.

A second organization in our sample reported dropping the use of an assessment center because of a reduction in business. There were not as many job openings as there had been when their business was growing; consequently, the assessment center was expensive and unnecessary for them to operate. The issue of the costs of an assessment center versus the benefits reaped for an organization is of great importance, though such cost/benefit analyses are often difficult to compute. We shall devote more attention to this issue in Section IV of this report.

Another reason given for dropping the use of an assessment center was the finding by an organization that the job description and skills evaluated in an existing center no longer converged with the skills and dimensions of the target job. It found that, over time, the job had changed sufficiently such that the skills being evaluated were no longer relevant to the decisions its people were trying to make. This experience clearly suggests that organizations must continually update their centers based on the changing nature of the organization and related jobs.

Another organization dropped its assessment center for a different set of reasons. The organization in question became concerned over the use and abuse of the voluminous assessment data. First of all, its assessment center was designed, intended, and presented to assessees as a developmental exercise. Nonetheless, organizational control of the data was poor, and allowed managers to obtain access to the assessment data on candidates for promotion. This violated the promises of confidentiality and privacy made to participants. Managers were in a position of sufficient power to be able to force disclosure of the data. At the same time, the assessment center staff was concerned over the match between (a) the organization's jobs and requisite skills, and (b) the situations, dimensions, and assessments derived from the prepackaged generalized assessment program they had in place. In short, this organization was so concerned about the appropriate use of its assessment data that it ultimately dropped the existing center. It remains committed to the concept, however, and is beginning efforts to design a more appropriate, controllable assessment process. This experience clearly points to the need for planning the uses of the data prior to center implementation. Many other organizations are also faced with this issue of data management, but have not been forced to discontinue the application.

These experiences point to one final conclusion regarding the consequences for an organization of implementing an assessment center. The effective design and implementation of an assessment center requires resources and a staff component: (a) to continually modify the procedures based upon research and experience; (b) to design measures of the relative cost-effectiveness of the process; and (c) to manage the use of the data within the organization. Without such expertise, the consequences for the organization may be far less than ideal.

Conclusions and Implications for Educational Applications

This review of the current state of the art in practice raises several practical issues for the reader who is seriously considering the use of an assessment center in an educational setting. We have noted, at several points, the need for further research on, or consistent standards for, many aspects of the design and operation of a center. Educators may or may not choose to become involved in such research; nevertheless, each step in the design and operation of an educational assessment center will demand careful consideration and reasoned choice among an array of alternatives. Let us summarize these issues briefly.

The number of applications of the assessment center technology in educational settings seems to be growing steadily. The reasons for these applications are as diverse as those cited by our sample from industrial settings. These applications may stem from a basic interest in educational innovation, or from a more specific interest in new methods for student or faculty appraisal. Many educators are now facing pressures from school boards, alumni/ae or the public to be more accountable for the outcomes of the educational process. Traditional outcome measures (e.g., grades or paper-and-pencil achievement tests) are coming to be viewed with some suspicion. The trend towards the use of behaviorally-based outcome measures, particularly those which can be conveniently translated into the world of work, seems inevitable. As a consequence, the assessment center technology may become more attractive to educators. We would simply caution the interested educator to be as clear as possible about the reasons underlying an application at the initial stages of consideration. Can your center become an adequate and cost-effective response to these underlying issues?

We have noted that assessment centers are currently being used for a variety of purposes, including selection and promotion of personnel; identification of long-range potential;

and diagnosis of training and development needs of personnel. Educational applications have also been directed to these purposes, though the trend in both industrial and educational settings appears to be towards increased use of the center as a diagnostic toal. Since the educational process usually involves evaluation, training and placement decisions at student, faculty and/or administrative levels, the assessment center technology may be able to provide much better performance-based data adaptable to each of these purposes. It would appear to be critical that the educator clearly articulate the purpose of the center at the early design stages, since this purpose will guide the choices to be made concerning nearly every other aspect of the design and operation of the center (e.g., the performance criteria, the types of dimensions to be measured, the nature of the feedback to assessees, or the methods for validating the assessments made in the center). We will return to this issue in the final section of this report.

The state of the art is probably somewhat more advanced for those applications aimed at employment decisions. Yet the more recent increase in developmental applications is perhaps more important to the educational community. We would speculate that the use of the assessment center technology as a diagnostic tool for student learning needs (and outcomes) will become the more common educational application in the coming decade. Thus, educators will need to be concerned with the special design and evaluation problems of this type of application.

For example, the reliability and validity of assessor judgments are obviously critical to employment-oriented applications. However, many practitioners in our sample--especially those involved with developmental centers--are not conducting reliability and validity research. It is our contention, shared with some writers in the field (e.g., Boehm & Hoyle, in Moses & Byham, 1977), that developmental applications may require even greater attention to the reliability and validity of the center, particularly the reliability and validity of the assessor ratings on specific behaviorial dimensions (as opposed to the overall competency judgments). It is likely that career guidance in educational settings will be much more focused upon these dimensions as representative of key strengths and weaknesses of the individual. Detailed feedback and interpretations on these dimensions for students must be based on reliable and valid assessments. Consequently, the educational community may have to invest some resources in the research necessary to base these dimensional judgments on sound measurement criteria.

Regardless of the purpose of the education-based center, its design and operation will certainly require a variety of internal and external resources. Many industrial organizations rely on a small number of consulting organizations for materials and/or design expertise. A few are fortunate to have the in-house expertise and staff to support the continuing development of the center. Clearly, education will have to devote some thought to the selection of external expertise for help, at the minimum, for the design of the center. The technology is far too complex to rely solely on internal resources. Obviously, the selection of these consultants should be based on their prior design experience (particularly in educational settings), their methodological sophistication, and their ability to produce materials and procedures with high utility and face validity for educational settings. This process may not be easy.

In addition, the educator will need to consider in advance the kind of internal support which will be required for the application. This would include such considerations as staffing for the administration of the center, the selection and training of assessor teams, the production of assessment materials, the management of the data generated, and/or the follow-up needed as a result of the assessment center process. These activities can and do require some degree of resource allocation and expenditure. These procedures, and their costs, must be reviewed carefully at the early stages of center design. Certain applications may present prohibitive resource expenditures; yet cost-cutting may seriously dilute the effectiveness of the process. This is undoubtedly one of the most difficult decisions facing the interested educator. We shall address this issue of cost/benefit analysis more directly in the next section of this report.

The creation of the behavioral dimensions to be measured in an assessment center can also present problems to the educational application. We have noted a great number of dimensions in use in practice, and an equal diversity in the manner by which these dimensions are derived. The standards of practice suggest that these dimensions should be derived from a rigorous and empirical analysis of target jobs. However, most educational applications will probably be oriented to a variety of jobs and careers; therefore, the use of a typical job analysis could be problematic. To date, educational applications have tended to rely on "armchair" speculation for the creation of dimensions, and have typically generated assessments on a much larger number of dimensions. It may well be necessary to measure assessees on a broad range of dimensions; however, it is reasonably clear that assessors may be faced with great difficulties in validly evaluating an assessee's performance across a great number of dimensions. This problem can be reduced by designing performance situations in which one assessor team measures only a few dimensions at one time or by spreading out the assessee's participation in the center over a longer period of time (this is just the approach taken by one educational institution in our sample). Or one could reduce the number of dimensions by designing rigorous job analyses of core job families towards which graduates are typically oriented (the costs here would be a problem). Or one could attempt to reduce the number of dimensions by measuring only those individual characteristics which appear (through research or theory) to be underlying causes of effective performance in a variety of job families. (This is the approach of the competency-based testing movement. The only caution here is that this approach represents somewhat of a return to the use of underlying constructs which would have to be validated through actual behavior.) In short, the step of creating behavioral dimensions for measurement in the center may prove to be an extremely difficult problem for many educational applications. The effectiveness of the application will surely depend on the rigor applied at this stage of assessment center design.

We found that in current practice most centers rely on a small number of fairly common performance situations to elicit assessee behavior (i.e., the in-basket, the leaderless group discussion, role-plays, and structured interviews). We also found that there are very few published standards for guiding the design or selection of these situations. Many users have modified the standard packages offered by consulting organizations in order to match their organizational context more closely. Given the purpose of the specific educational application, educators will need to insure that the situations are truly representative of the demands of the target job or jobs. For example, the use of an in-basket or a leaderless group discussion may be quite relevant to selection or developmental diagnosis for administrative positions; it would probably be irrelevant to selecting faculty for skills needed in the classroom. The educator may need to employ both internal and external experts to help design these situations to elicit most effectively the types of behavior which are critical to performance effectiveness. Lack of attention to this step will certainly be very costly in the long run.

We have noted that the common practice in composing assessor teams is to utilize supervisors one or two levels above the assessee. The different goals and organizational structure of educational institutions may necessitate the use of different combinations of assessors. In the case of selection for faculty or administrative postions, the assessor teams might be composed of present faculty or senior administrators, which would be congruent with current practice in industrial settings. However, the use of an assessment center for student appraisal may require assessors other than faculty members, since faculty may not possess any special knowledge of the jobs for which assessees are preparing. One alternative is to employ job experts from outside organizations, an approach effectively utilized by a few different educational users. In any case, the educator will need to compose these assessor teams to maximize their credibility in the eyes of the assessees, and will need to consider the procedures and costs involved in soliciting assessor team members. We suspect that these are not easy problems to solve in educational settings.

We have noted that the quality of assessor training is critical to the quality of assessor judgments, and therefore a fundamental determinant of the effectiveness of the center. In our survey of the state of practice, we found much initial commitment to assessor training. Unfortunately, that commitment too often dwindles over time and few organizations do any rigorous certification or retraining of assessors. Within educational applications we found the state of the practice to be equally problematic if not worse. The educator may be tempted to assume that assessors will be good evaluators of performance because faculty and administrative personnel conduct evaluations as a routine part of their jobs. This is a dangerous assumption which will likely severely limit the validity of the application. The focus of the assessor's observations will be different from typical educational evaluations; therefore, assessors will require effective training. This may increase some resistance from faculty or administrative personnel, particularly if the educator needs to rely on external expertise for the design or conduct of the training. Nevertheless, this is a critical step in implementation and simply cannot be overlooked or avoided.

Assessees being evaluated in the assessment center may come to participate in a variety of ways. Current practice varies considerably with respect to whether participation is voluntary or mandatory, or whether the assessee nominates himself or herself versus being nominated by a supervisor. The basic purpose of the application may well dictate this choice. For student assessment, it may well be desirable to have mandatory participation, with some free choice as to when the student participates. In contrast, the institution may wish to offer the assessment as a voluntary, or adjunct, service as part of career counseling. For faculty or administrative assessment, the number of variations is perhaps far greater. In any case, the educator must recognize the potential influences of the assessee selection process on the attitudes, expectations and resultant motivation of the groups to be assessed, particularly when the entire process represents a significant departure from past evaluation practices. Effective management of this transition is clearly fundamental to the success of the implementation effort.

The assessor observations, dimensional ratings and overall competency judgments are the basic products of the assessment center process. We have found a rather great diversity, in current practice, in the manner in which these products are created. The educator will need to make reasoned choices about: (1) the methods by which observations of behavior are recorded; (2) the types of rating scales to be employed on dimensional ratings (including the number of scale points or the degree of inference allowed in these ratings); and (3) the decision rules for summarizing multiple assessor judgments on a specific dimension, or for obtaining (when desired) an overall competency judgment based on summarizing the dimensional ratings. There is a paucity of research to guide these decisions; however, the choices should reflect the kind of data needed as a result of the assessment and the uses to which the data will be put. For example, employment decisions will likely require some overall competency judgment from the assessor team; in contrast, developmental applications will more likely be focused on the specific dimensional ratings. Furthermore, any ratings should be thoroughly grounded in observable behaviors and rest upon decision rules which are as explicit as possible. Lack of attention to these issues will seriously limit the acceptability of the assessment to assesses and cast undue suspicion on the validity of the results.

The feedback of assessment center results in some form to assessees is a fairly standard practice. This will certainly be an important concern in future educational applications. The educator will need to decide upon the format of this feedback, keeping in mind the resources required to produce oral or written reports and/or to conduct follow-up feedback or counseling sessions. These practices can represent relatively hidden, but necessary, resource expenditures which need to be actively planned from the initial design of the center. In addition, educators will need to decide (given the purpose of the center, of course) whether any data will be retained by the institution and how these data will be stored. Issues of the confidentiality of the data, and their relationship to other institutional records of performance, are surely important topics for consideration very early in the implementation process.

Finally, we have noted several commonly cited and/or possible consequences of implementing an assessment center for both participants and the organization. Both types of consequences require careful examination by the educator prior to implementation. With respect to the individual consequences, the educator must be concerned with the nature of the individual career decisions which will be made on the basis of the assessment results. Will these decisions be made on the basis of the assessment results? Will these decisions be based on reliable and valid information that is of added value to present appraisal or counseling practices? Can the impacts on individuals be examined over time?

Perhaps a more critical issue for the educator is the potential consequences of assessment center implementation for the institution per se. With the possible exception of voluntary, developmentally-oriented centers, the implementation of an assessment center will likely represent a significant departure from current appraisal practices in most educational institutions. Educators are likely to experience at least some resistance from all groups in the institution affected by the change. Students may resist the "changes in the rules" when the center is used for measuring academic accomplishment. Faculty may resist the challenge to their traditional prerogative of evaluating students within the confines of their specific discipline and may resent the additional duties and responsibilities necessary as a member of an assessor team. Administrators may be hesitant to become involved in the potential conflict which might arise as a result of these changes, and may wish to avoid the headaches associated with the rigorous administration of a center. The educator will clearly have to involve all of these groups in any implementation in order to build the consensus necessary for implementation. This is undoubtedly a difficult and time-consuming task.

In summary, this review of the current state of practice of assessment center technology provides ample demonstration that the concept is both technically sophisticated and operationally complex. It is not a technology to be blindly utilized by the novice. Nor is it a technology that should be embraced as the ultimate assessment technique. Current practice is best characterized as a significant advance over previous, and more simple, techniques; yet the state of the art is confronted with many compelling problems requiring research-based standards. In too many instances, current practice falls short of the ideal. The educator's awareness of these practical, and complex, issues should increase the probability of successful implementation of the technology in future applications. Conversely, lack of attention to these issues will only dilute the technical advance represented in the concept and will likely create a significant probability of failure.

TABLE 7A

Working List of Assessment Center Users and Designers

Aerospace (2)

Lockheed Missiles & Space Co.
Martin Marietta Corporation

Airlines (5)

American Airlines
British Airways Corporation
Eastern Air Lines
Western Air Lines
Qantas Airways, Ltd.

Appliances (2)

McGraw-Edison Company
Northern Telecom Inc.

Assessment Package Designers, Suppliers, Cooperatives (16)

American Management Assoc. (AMA)
Assessment Designs, Inc. (ADI)
Center for Creative Leadership (CCL)
Consulting Associates
Development Dimensions, Inc. (DDI)
Educational Research Corp. (ERC)
Educational Testing Service (ETS)
Great Lakes Assessment Council (GLAC)
Institute for Personality Assessment Research
Life Insurance Marketing & Research Association (LIMRA)
Life Office Management Assoc.
Personnel Decisions, Inc. (PDI)
Selection Consulting Center (SCC)
Syracus Research Corporation
U. of Texas, Adult Performance Level Project
Zenger & Miller Associates (formerly with Syntex)

Automotive (3)

Ford Motor Co.
General Motors Corporation
Chrysler Corporation

Banks & Holding Companies (6)

Bankers Trust New York Corp.
Bank of New York Co.
Federal Reserve Bank of Chicago
Federal Reserve Bank of New York
Maryland National Bank
Virginia National Bank

Chemicals (4)

Mallinckrodt Inc.
Monsanto Co.
Stauffer Chemical Co.
Union Carbide Corp.

Containers (1)

Owens-Illinois, Inc.

Drugs (13)

Ciba-Giegy Corp.
Eli Lilly (Germany)
Hoffman-LaRoche, Inc.
Jack Eckhardt Drug (retail)
Merck, Sharp & Dohme
Ortho Pharmaceutical Corporation
Pfizer Company (Canada)
Richardson-Merell Inc.
S.C. Johnson & Son Inc.
Schering Corp.
Schering-Plough Corp.
Smithkline Corporation
The Upjohn Company

Education: Postsecondary ([illegible]7)

Alverno College (WI)
Baylor University (TX)
Boston University, Women's Leadership Project
Brigham Young University, Graduate School of Business

Table 7A cont.

Education: Postsecondary cont.

Colorado State University
Eastern Michigan
Elmhurst College (IL)
Howard Universtiy, Institute for Minority Business Education (Wash, D.C.)
Kalamazoo College (MI)
Mission College, San Jose (CA)
National College of Education, Chicago
Pace University (NY)
St. Mary's College (CA)
Stanford University, Graduate School of Business (CA)
University of Phoenix (AZ)
University of Redlands (CA)
Nova University (FL)

Education: Secondary (22)

Broward County, Florida Public School Teachers
Greensboro, North Carolina Public Schools
Merrimack Education Center Collaborative (MA)
National Academy of School Executives
National Association of Secondary School Principals
*New England Association of School Business Officials
New York State Hebrew Day School Executives
*6 New York State Regions
*Norwalk, Connecticut
Pinellas County School System (FL)
Portland, Oregon
Rhode Island State Dept. of Education
So. Orange, New Jersey Public Schools
*State of Maryland
*State of Wisconsin
*12 Texas Regional High School Programs
Title VII Bilingual Project Directors (New York City, Boston, San Francisco)
U. of Nevada, Dept. of Sec. Ed. (designed for H.S. Diploma granting)
Wisconsin State Education Assoc.
Woodbridge, New Jersey

Electrical, Electronics (7)

General Dynamics Corp. (Electrical Boat Division)
General Electric Co.
Harris Semi-Conductor Corp.
Motorola Inc.
RCA Corporation
Varian Associates
Westinghouse Electric Corp.

Food and Lodging (2)

Harrah's Tahoe

Government: Federal (34)

Public Service Commission (of Canada)
Bureau of National Affairs
Internal Revenue Service
U.S. Civil Service Commission
U.S. State Department, Foreign Service Officers
Federal Aviation Administration
U.S. Postal Service
U.S. Forest Service
Federal Bureau of Investigation
Public Service Board (Australia)
U.S. Dept. of Interior
U.S. Dept. of Agriculture
USDA Animal and Plant Health Inspection Service
U.S. Dept. of Commerce
U.S. Dept. of Treasury, Customs Service
Bureau of Engraving and Printing
U.S. Marshalls Dept. of Justice
Social Security Program
Food and Drug Administration

---

*external high school diploma program

Table 7A cont.

Government: Federal cont.

U.S. Dept. of Housing and Urban Development
State Dept.
U.S. Office of Education
U.S. National Parks Service
U.S. National Highway Traffic Safety Commission
U.S. General Accounting Office
U.S. Dept. of the Interior
U.S. Dept. of Labor
Bermuda Federal Government
Bureau of Alcohol, Tobacco and Firearms
State Organization for Administration and Employment Affairs (Iran)
U.S. Dept. of Health, Education and Welfare
U.S. Office of Management and Budget

Government: Local (16)

Akron, Ohio (fire)
City of Detroit
City of Omaha
City of Chicago
Clearwater, Florida (fire)
Flint, Michigan
City of Houston
Kansas City Police/Fire
New York City Police
Oklahoma City and City Fire Dept.
St. Louis County Police
St. Louis Police
St. Paul Companies
St. Petersburg Police
Richmond Police/Fire
Rochester Police/Fire

Government: State/Regional (13)

Prince Edward Island Civil Service Commission
State of Iowa, Dept. of Public Safety
Government of Quebec
Government of Manitoba
Justice Institute of British Columbia, Police Academy
Hennepin County, Minnesota
California State Personnel Board
State of Illinois
State of Colorado
State of Wisconsin
State of Michigan
State of Nebraska
Orange County, California

Instruments (Measuring and Scientific), Photographic Equipment (3)

American Hospital Supply Corp.
Eastman Kodak Company
Itek Corp.

Metals and Mining (2)

Aluminum Company of Canada Ltd.
Kaiser Aluminum & Chemical Corp.

Military (3)

Air Force
Army
Navy

Natural Resources (9)

Agway, Inc.
Atlantic Richfield Company
Continental Oil Co.
Exxon Corporation
Mobil Oil Corporation
Shell Oil Co.
Syncrude, Canada, Ltd.
Tenneco, Inc.
Texaco, Inc.

Non-bank Financial (3)

Fireman's Fund American Insurance Companies
SAFECO Insurance Company
South African Mutual Life Assurance Society

Table 7A cont.

Office Equipment, Computers (10)

Control Data Corporation
Data General Corporation
GTE Data Services
International Business Machines Corp.
NCR Corporation
Pitney-Bowes Inc.
Rank-Xerox of South Africa
Steelcase, Inc.
Xerox Corporation
York Graphics Services, Inc.
Texas Instruments Inc.

Paper & Forest Products (1)

Boise Cascade Corporation

Personal Care Products (2)

Bristol-Myers Company
The Procter and Gamble Company

Publishing (3)

Knight-Ridder Newspapers Inc.
McGraw-Hill Book Company
The New York Times Co.

Railroads (1)

Southern Railway Company

Retailing (Food) (3)

Gino's, Inc.
McCormick & Company, Inc.
Supermarkets General Corp.
Domino's Pizza

Retailing (Non-food) (6)

Foley's
J.C. Penney
Sears, Roebuck Co.
Steinberg's Properties Ltd.
Target Sportswear, Inc.
Wickes

Savings & Loan (1)

Florida Federal Savings & Loan Assoc.

Service Industries (12)

Aid Assoc. for Lutherans
AAA (Michigan) American Automobile Assoc.
Blue-Am/Blue Shield (Boston)
Blue-Am/Blue Shield (Jacksonville)
Blue Cross/Blue Shield Assoc. of Ill.
Chelsea CETA Consortium
Hospital Assessment Center
Northside Mental Health Center
The Ralph W. Parsons Co.
University of Michigan Hospital
Washington Transit Authority

Special Machinery (5)

Caterpillar Tractor Co.
Clark Equipment Co.
Dresser Industries Inc.
International Harvester Co.
Sundstrand Corporation

Steel (3)

Armco (Steel Corp.)
Florida Steel Corporation
United States Steel Corporation

Textiles & Apparel (1)

Hart Schaffner & Marx

Tire and Rubber (3)

The B.F. Goodrich Co.
The Firestone Tire & Rubber Co.
The Goodyear Tire and Rubber Co.

Tobacco (2)

Brown and Williamson Tobacco Corp.
Philip Morris Incorporated

Table 7A cont.

Trucking (1)

Questor Corporation

Utilities (20)

American Telephone and Telegraph Co.
Baltimore Gas & Electric Co.
Central Telephone & Utilities Corp.
Continental Telephone Corp.
Consolidated Edison Company of New York Inc.
Consumers Power Company
The Detroit Edison Company
Duke Power Company
Florida Power Corp.
General Telephone Co. of Florida
General Telephone Co. of Illinois
General Telephone Co. of Indiana
General Telephone Co. of Ohio
Manitoba Telephone Systems
Michigan Bell Telephone Company
Northern Illinois Gas Company
Ontario Hydro Corp.
Southern Bell Telephone & Telegraph Co.
Tampa Electric Co.
Tennessee Valley Authority

Miscellaneous/Other (16)

Cydsa SA (Mexico)
Edgars, Ltd.
Federal Express Corp.
Foodplex, Inc.
The Heil Co.
Huyck Corp.
Jet Propulsion Laboratories
ICI Americas Inc.
Industrias Resistol SA (Mexico)
Institute for New Enterprise Development
Irving Trust Co.
Kendall Company
Negromex, SA (Mexico)
Oak Ridge National Laboratories
Southwire, Co.
USM Corporation

TOTAL N = 274

TABLE 7B

Sample Distribution by Organization Type and Quality of Data

ORGANIZATION TYPE	CONTACT OR COMPLETE DATA	PARTIAL DATA	IDENTIFIED ONLY	TOTAL
Assessment Suppliers				
Assessment Package Designers, Suppliers, Cooperatives	12		4	
	12	0	4	16
Consumer Products				
Beverages				
Drugs	1	1	11	
Food Processing				
Personal Care Products	1		1	
Tobacco			2	
	2	1	14	17
Consumer Sales				
Food and Lodging			2	
Leisure Time Industries				
Retailing (food)			3	
Retailing (non-food)	2		4	
Service Industries	3		9	
	5	0	18	23
Education				
Postsecondary	7	1	9	
Secondary	5	9	8	
	12	10	17	39
Financial				
Banks and Bank Holding Companies			6	
Non-bank Financial			3	
Real Estate and Housing				
Savings and Loan			1	
	0	0	10	10

Table 7B cont.

ORGANIZATION TYPE	CONTACT OR COMPLETE DATA	PARTIAL DATA	IDENTIFIED ONLY	TOTAL
Government				
Local	3		13	
State and Regional	5		8	
Federal	2		32	
Military	2		1	
	12	0	54	66
Manufacturing				
Appliances			2	
Automotive	2		1	
General Machinery				
Miscellaneous Manufacturing			1	
Special Machinery	1		4	
	3	0	8	11
Media				
Publishing			3	
Radio and TV Broadcasting				
	0	0	3	3
Technical				
Aerospace			2	
Electrical, Electronics			6	
Instruments			3	
Office Equipment, Computers	2		9	
	2	0	20	22
Resource Suppliers				
Chemicals	1		3	
Metals and Mining			2	
Natural Resources (fuel)	3		6	
Oil Service and Supply				
Paper and Forest Products			1	
Steel			3	
Tire and Rubber			3	
Utilities	2		18	
	6	0	36	42

Table 7B cont.

ORGANIZATION TYPE	CONTACT OR COMPLETE DATA	PARTIAL DATA	IDENTIFIED ONLY	TOTAL
Secondary Processing				
Building Materials			1	
Containers				
Textiles and Apparel			1	
	0	0	2	2
Transportation				
Airlines	1	1	3	
Railroads			1	
Trucking			1	
	1	1	5	7
Miscellaneous				
	0	0	16	16
SUPPLIERS	12	0	4	16
INDUSTRY TOTAL	19	2	132	153
GOVERNMENT TOTAL	12	0	54	66
EDUCATION TOTAL	12	10	17	39
SAMPLE TOTAL	55	12	207	274

TABLE 7C

Target Jobs

The following list of job titles for target positions of assessment centers is designed to convey the diversity of applications of the assessment center technology. The numbers in the righthand column reflect the absolute frequency with which each job title has been the focus of users. The total N for this frequency count reflects all those centers for which the target position was known, not just those who directly shared data with the researchers.

JOB TITLE	TIMES MENTIONED
Education	
Administrative (college)	4
Adults without high school diplomas	18
Counselor teachers	1
Graduate students	5
Lecturers	1
Pharmacy students	1
School principals (secondary)	2
Undergraduate students	2
Women administrators	1
Government and Public Sector	
Agency for International Development (overseas managers)	1
Fire Department (assistant chiefs)	1
Police (entry level, lieutenant and captain)	1
Military (NCO's)	1
Military (officers)	1
Public Sector (managers)	2
Urban Planners (municipal employees)	1
Management	
Executives and upper level managers	7
Middle managers	7
Entry or first level managers	21
District or branch managers	4
Miscellaneous management titles	6
Miscellaneous Jobs and Titles	
Sales agents or representatives	2
Clerical	1
Instructors	1
Inspectors	1
Leaders from organizations of all types (industry, education, etc.)	2
Skilled laborers (non-management)	1
Paraprofessionals in mental health	1
Professional workers	1
	N = 99

TABLE 8

Reasons Cited by Users for Setting Up their Assessment Center(s)

Reasons	Absolute Frequency of Response
Dissatisfaction with tradition selection and promotion	*
Improved degree of managerial candidate	**
Anticipated strength of assessment vs. courts & EEO	**
Improving employment and personnel selection	3
To identify and develop leadership skills (provide feedback and learn to change)	2
Want to identify skills of underutilized people in dead end jobs	2
Minority involvement	4
Affirmative Action needs	2
Equal Employment Opportunity (EEO)	1
Evaluate skills of individuals for (H.S.) degree certification in nontraditional settings (outside school)	2
Wanted to assess individuals on job competence which isn't paper and pencil testable (i.e., interpersonal teaching skills)	1
Chronic turnover problem (seek selection solution)	2
Costs of turnover	1
Face validity important	2
Design a package to be used in-house without external involvement	1
Inquiries from clients about technique (after Civil Rights Act of 1964)	1
Had been considering designing diagnostic development program	1

* Twenty-one companies within one industry cited this as the most common reason for setting up an Assessment Center

** Frequently cited reason in same survey

Table 8, cont.

Reasons	Absolute Frequency of Response
Technology is a flexible technology with multiple applications to client idiosyncracy	1
Improve skills of people on job (who can't afford to quit and go to school full-time to get skills)	1
Saw management problems as boiling down to a need for career planning and development	1
Saw ass't as way to credit test shy people for skills	1
Wanted to be sure award of certificate reflects skill	1
Pilot to see if method would help select management	1
Method to identify skills and determine training and development for internal job applicants for upper level jobs (mgt)	1
Wanted a multiple hurdle process where organization or applicant would bow out of selection process	1
Need to move people up quickly due to growth in business	1
Needed to develop people rapidly and to identify those who are already prepared	1
Returned to academia from industry; was distressed at woeful lack of meaningful management training	1
Felt schools needed to give better feedback on appropriateness of career choice	1
Wanted to be on cutting edge of selection methodology	1
Determine whether students meet the training standards, and also whether curriculum is providing adequate training for students	1
Response to public and to police contention over police department promotion system	1
Provide diagnostic and career planning information about alternate career paths	1
Need to assess and develop peer leadership skills	1
Demonstration of ass't methodology applicably to educational settings	1
Get on bandwagon - everyone using it	1
Upper management strongly committed to it	1
Wanted nontraditional skills among teachers	1
Use in related part of the organization	1

Table 8, cont.

Reasons	Absolute Frequency of Response
To develop individuals already identified and selected for job (9-18 months in future)	1
Need to predict project manager skills	1
Find teaching skills not accessible with paper and pencil	1
To identify possible changes in career path for people in dead end jobs	1
TOTAL	51

TABLE 9

Resources Used to Design and Implement Assessment Centers

ORGANIZATION TYPE	NATURE OF IN-HOUSE INVOLVEMENT	USE OF MAJOR SUPPLIER	USE OF OTHER CONSULTANTS/SERVICES
(10)* Suppliers	(10) in-house design		(2) consultant brought in who had worked with them before; most suppliers relied on client support services for ideas in design phase
(3) Consumer Sales	(1) in-house staff discussed and validated process (2) totally in-house design		(1) design done by custom design consultant group
(5) Financial	(3) in-house design	(2) used major suppliers	
(2) Resource Suppliers	(1) in-house design	(1) used a major supplier	
(1) Consumer Products	(1) in-house design		
(1) Transportation	(1) in-house support	(1) used a major supplier	
(2) Technical	(2) major in-house support	(1) used a major supplier's materials	(1) used external academic consultant

* Number of organizations supplying useable, detailed data

Table 9 cont.

ORGANIZATION TYPE	NATURE OF IN-HOUSE INVOLVEMENT	USE OF MAJOR SUPPLIER	USE OF OTHER CONSULTANTS/SERVICES
(3) Manufacturing	(3) rarely in-house design		(3) used unspecified external consultants
(11) Government			
Local	(2) in-house design	(2) used a major supplier	
State	(2) totally in-house design	(1) used a major supplier	
Federal Agencies	(3) major in-house support (e.g., validation)	(2) design expertise from a major supplier; modified materials	
Military	(1) primarily in-house design	(1) modified supplier's original materials	
(8) Education			
Secondary	(1) no in-house design		(3) used an outside consultant group with educational expertise or a group especially convened to adapt industry experience for an educational project
Postsecondary	(2) primarily in-house design	(2) relied primarily on major suppliers for materials or modified materials minimally	

TABLE 10

Purposes of the Assessment Center

ORGANIZATION TYPE	PURPOSE				CRITERION IF ARTICULATED	VALIDATION
	# cases	Select	Promote	Develop		
(12) Suppliers	3	✓	✓	✓	varies by client	up to client
	2	✓		✓	life skills; life skills and individual skills	no
	3			✓	(2) develop leadership skills	no
					(1) job skills as identified by job analysis	yes
	2		✓		unclear	unclear
	2	UU*	UU	✓	performance in managerial job	yes
(3) Consumer Sales	1	✓			personnel director's rating; self-rating; immediate supervisors' ratings; personnel records	yes
	1		✓	✓	unclear	content validity
	1			✓	job skills identified for target job	no
(6) Financial	2			✓	unclear	(1) in progress (1) yes
	1	✓			turnover; success in sales; move into management	in progress

* Unintended use

Table 10 cont.

ORGANIZATION TYPE	PURPOSE				CRITERION IF ARTICULATED	VALIDATION
	# cases	Select	Promote	Develop		
Financial cont.	1	✓		✓	pilot stage; unclear	in progress
	1	✓	✓	UU	unclear	in progress
	1	✓	✓		unclear	yes
(2) Resource Suppliers	1	✓			unclear	seems content valid
	1		✓	✓	job success (e.g., salary advancement, merit rating)	yes
(1) Consumer Products	1	UU		✓	unclear	yes; in progress
(1) Transportation	5	✓			unclear	no
(2) Technical	1	✓	✓		turnover	no
	1			✓	success as manager (measure unclear)	no
(3) Manufacturing	3	✓	✓		unclear	yes
(11) Government						
(4) Local	1	✓			pass probationary period on job; turnover	in progress
	1		✓		unclear	content valid according to supplier

Table 10 cont.

ORGANIZATION TYPE	PURPOSE				CRITERION IF ARTICULATED	VALIDATION
	# cases	Select	Promote	Develop		
Local Gov't cont.	1	✓	✓		unclear	in progress
	1	✓		UU	unclear	unclear
(3) State	1		✓		unclear	yes
	1	✓			turnover	yes
	1		✓	✓	unclear	content valid according to supplier
(3) Federal Agencies	1	✓			unclear	yes
	2		✓	✓	unclear	yes, also in progress
(1) Military	1			✓	unclear	yes
(8) Education						
(3) Secondary	1	✓			standard principal job description	no
	1		✓	✓	unclear	no
	1			✓	unclear	no
(5) Post-secondary	4			✓	(3) generic notion of competence for a class of jobs	no
					(1) specific job analysis results for a target job	yes
	1	✓			student evaluation of faculty	no

TABLE 11

Number and Sources of Behavioral Dimensions in Assessment Centers

ORGANIZATION TYPE	# DIMENSIONS	METHOD OF DIMENSION GENERATION
Suppliers		
Educational assessment center is product	23	Brainstormed with renowned educational leader; job analysis by means of observation
	64 in 5 groups	Panels of community members, educators, people with different jobs met to identify themes
	65 in 41 modules	Not given
Assessment centers for diverse jobs and/or organizations	25-30	Ranged from brainstorming to job analysis (depending on client and consulting content)
	17 12	Supplier's staff generated list of ideas, surveyed clients, sorted survey and staff ideas, and developed a generic list for each of two different centers
	11	Not given
	11	Job analysis, interviews, workshops, questionnaires, sorted ideas
	8-10	Interviews with incumbents generally focused on tasks
	Not given	Job analysis
	8	Not given
Consumer Sales	Not given	Job analysis
	11	Job analysis of unspecified form
Financial	8	Job analysis using interviews and panels with incumbents and managers

Table 11 cont.

ORGANIZATION TYPE	# DIMENSIONS	METHOD OF DIMENSION GENERATION
Financial cont.	14	Job analysis (form unspecified)
	12	
	11	
Resource Suppliers	Not given	From supplier files
	9	Job analysis (form unspecified)
Consumer Products	7	Job analysis, open-ended discussion and critical incidents from incumbents
	7	
	Not given	
Transportation	10	Job analysis including discussion with incumbents, mailing of panel lists to be rated by supervisors and managers
	10	
	10	
	12	
	16	
Technical	12	Job analysis, meetings with upper management and regional staff, plus observation of branch operations
Manufacturing	8	In-house research, plus work with other consultants
	8	Job analysis (form unspecified)
	10	Not given
Government		
Local	8	Job analysis (form unspecified)
	10	Job analysis, observation, interviews
	12	Job analysis (primarily interviewing)
	13	Job analysis (brainstorming with panel of experts)
State	6	Not given
	11	Job analysis (form unspecified)
	16	Provided by supplier, apparently from files

Table 11 cont.

ORGANIZATION TYPE	# DIMENSIONS	METHOD OF DIMENSION GENERATION
Federal Agencies	9	Free-response form to incumbents to rank most relevant characteristics (top nine kept)
	7	Job analysis (unspecified form)
	6	Armchair identification of generic skills for variety of jobs (all targets of one center)
Education		
Secondary	23	From supplier package
	12	Job study using job description file information on performance appraisal interview with incumbents
	11	From supplier package
	16	Apparently from supplier's files
Postsecondary	Not given	Academic audit of faculty by student evaluations
	17	Informal job analysis with a small number of incumbents
	10	Pulled from files of two major suppliers and from published research
	20	Borrowed from research and published supplier lists

TABLE 12

Behavioral Dimensions Utilized in Assessment Centers

DIMENSION	ABSOLUTE FREQUENCY OF USAGE	DIMENSION	ABSOLUTE FREQUENCY OF USAGE
Ability to Analyze and Solve Problems	23	Oral Presentation Skill	4
Planning & Organizing Skill	23	Persuasiveness	4
Resistance to Stress	21	Planning	4
Interpersonal Sensitivity Skills	21	Range of Interests	3
Oral Communication	21	Independence (Self-Sufficiency)	3
Leadership	19	Works for Teamwork	3
Written Communication	17	Sees that Established Goals are Achieved	3
Controlling (Management Control)	15	Human Relations	3
Decision-Making	15	Functional	2
Initiative (Self-Starting)	15	Organizational Orientation	2
Decisiveness	15	Thoroughness	2
Flexibility	13	Budget Analysis Skills	2
Persistence (Tenacity, Perseverance)	12	Motivated Others	2
Delegation	9	Genuineness	2
Judgment	9	Educational Perspective (Values)	2
Organizing	8	Ability to Learn	2
Creativity	8	Integrity (Honesty with Others)	2
Risk-Taking	7	Development of Subordinates	2
Accountability	6	Organizational Sensitivity	2
Work Standards	6	Self-Motivated	2
Impact	6	Takes Responsibility for Performance	2
Energy	5	Gives Feedback on Performance	2
Listening Skill	5	Social Perception (Sensitivity)	2
Communication Ability	5	Person Perception	2
Empathy (Sensitivity)	5	Achievement Drive	2
Perception	4	Attitude Toward Self	2
Motivation for Work	4	Administrative Skills	2

Table 12 cont.

DIMENSION	ABSOLUTE FREQUENCY OF USAGE	DIMENSION	ABSOLUTE FREQUENCY OF USAGE
Activity Level	1	Self-Structure	1
Led the Discussion	1	Analytical Skills	1
Influenced Others	1	Handle Uncertainty	1
Task Orientation	1	Handling & Coping with Detail	1
Self-Disclosure	1	Independence/Assertiveness	1
Respect	1	Teaching & Supervising Others	1
Political Behavior	1	Anxiety Level	1
Group Leadership	1	Concreteness of Decision	1
Individual Leadership	1	Ability to Establish Rapport	1
Considerateness	1	Research Skills	1
Career Ambition	1	Isolation	1
Resilience	1	Responsiveness	1
Technical Translation	1	Poise	1
Political Sensitivity	1	Interaction with Subordinates	1
Sets High Achievable Standards for Performance	1	Interaction with Peers	1
Provides Technical Leadership	1	Organizational Acumen (Political Sensitivity)	1
Educational Level	1	Process-Oriented	1
Information-Gathering	1	Forcefulness	1
Follows Through on Details	1	Doesn't Back Down (Not Easily Threatened by Adult Learners)	1
Behavioral and Emotional Maturity	1	Willingness to Lead	1
Professional Attitude Toward Self	1	Staffing	1
Working with People	1	Adherence to Duty and Responsibility	1
Capacity for Delay	1	Ethical Responsibility	1
Objectivity	1	Control and Follow-up	1
Organization	1		

TABLE 13

Types of Assessment Techniques Utilized in Assessment Centers

ASSESSMENT TECHNIQUE	ABSOLUTE FREQUENCY OF USAGE	DESIGNED FOR PARTICULAR ORGANIZATION AND JOB	DESIGNED BY SUPPLIER FOR SIMILAR JOBS
Cooperative group discussion (leaderless)	43	36	7
In-Basket	39	17	22
Competitive group discussion (leaderless)	37	9	28
Role Play	21	21	0
Structured interview (background and career-related)	20	12	8
Interview role-play	14	4	10
Business game	13	5	8
Written presentation	12	4	8
Oral presentation	12	6	6
Case study	10	6	4
Written exercise problem-solving	8	2	6
Paper-and-pencil tests	7	0	7
Scheduling exercise	7	1	6
Psychological tests (aptitudes, personality, projective tests)	6	0	6
Background form (biodata or career path)	6	0	6
In-Basket interview	6	1	5
Assigned role cooperative discussion	5	0	5
Analysis and critique of someone else's performance	4	3	1
Outdoor exercises (USAF)	3	3	0
Observation of unstructured social situation (e.g., luncheon)	3	1	2

Table 13 cont.

ASSESSMENT TECHNIQUE	ABSOLUTE FREQUENCY OF USAGE	DESIGNED FOR PARTICULAR ORGANIZATION AND JOB	DESIGNED BY SUPPLIER FOR SIMILAR JOBS
Management questionnaire (what would you do if...hypothetical situations)	3	0	3
Assigned role competitive discussion	3	0	3
Supervisor ratings	2	0	2
Film case	2	0	2
Peer rankings	2	0	2
Letter-writing exercise	1	0	1
Panel interview	1	0	1
Peer interviews	1	0	1
Job-related activities questionnaire (e.g., education, reading, etc.)	1	0	1
Work diary (document past work activities, rationale)	1	0	1
Oral test	1	0	1
Self-Ratings	1	0	1
Interest questionnaire	1	0	1
Training exercise (act as trainer)	1	1	0
Reorganization exercise	1	1	0
Business planning exercise	1	1	0

TABLE 14

Sources for Assessor Team Composition

ORGANIZATION TYPE	SUPERVISORS Up 1 level	SUPERVISORS Up 2 levels	SUPERVISORS Level Unspecified	PEERS	PSYCHOLOGISTS	OTHER
(11) Suppliers			(6)		(2) have their staff of psychologists act as assessors	(1) used a group of supervisors from a variety of organizations that the mixed assessee group comes from (2) used teachers as assessors
(3) Consumer Sales			(2)			(1) central office personnel staff
(6) Financial		(1)*				
		(1)				(1) not given
			(3)			
(2) Resource Suppliers			(1)			
		(1)				
(1) Consumer Products			(1)			
(5) Transportation			(5)			
(2) Technical			(1)			
(3) Manufacturing			(1)			(1) not specified
		(1)				

* Cited by (1) cooperative research organization as norm for study of 21 firms

Table 14 cont.

ORGANIZATION TYPE	SUPERVISORS			PEERS	PSYCHOLOGISTS	OTHER
	Up 1 level	Up 2 levels	Level Unspecified			
(11) Government						
(4) Local			(1)			(1) supervisors from other municipal departments
		(1)				(1) unclear in pilot stage
(3) State			(2)			(1) agency heads from other areas (not in area of target jobs)
(3) Federal Agencies						(3) used mixed assessor pools of people outside agency, insiders, different jobs
(1) Military		(1)				
(8) Education						
(3) Secondary	(1)				(1) psychologist staff of consultant group	
			(1)			
(5) Postsecondary				(1) peers' faculty assess faculty		(3) faculty judge students
						(1) peers, faculty and outsiders assess students

TABLE 15

Assessor Training Methods
(Sample Frequencies of Method Use by Organization Type)

ORGANIZATION TYPE	Certification	Refresher Course Offered	Length	Role Play	Do Assessment Exercises	Read Material	View Tapes	Observe and Rate Behavior	No Structured Training	Not Specified
(11) Suppliers	1	3	(1) 20 hrs (1) 24 hrs (1) 1 day (2) 5 days (1) 3-5 days	5	5	6	5	5	1	5
(3) Consumer Sales	1		(1) 5 days		1	1		1		2
(5) Financial										5
(1)*			mode 16-20 hrs	✓	✓	✓		✓		
(2) Resource Suppliers		1	(2) 5 days	2	2	2	1	2		
(1) Consumer Products			(1) 4 days	1	1	1	1	1		
(1) Transportation			(1) 5 days	1	1	1	1	1		
(2) Technical			(1) 4-5 days						1	1
(3) Manufacturing	1		(1) 5 days (2) 2 days		1	2	1	2		1

* Cited by (1) cooperative research organization as the norm in a sample of 21 firms

Table 15 cont.

ORGANIZATION TYPE	Certification	Refresher Course Offered	Length	Role Play	Do Assessment Exercises	Read Material	View Tapes	Observe and Rate Behavior	No Structured Training	Not Specified
(11) Government										
(4) Local			(1) 2 days	2	2	2	1	2		2
			(1) unclear							
(3) State			(1) 2 days	3	3	3	2	3		
			(1) 3 days							
			(1) 5 days							
(3) Federal Agencies			(2) 5 days	2	2	3	2	2		1
(1) Military			(1) 5 days	1	1	1	1	1		
(8) Education										
(3) Secondary			(1) 5 days	1	1	1	1	1		2
(5) Postsecondary			(1) unclear	2	2	2		2	1	2
			(1) 4 hrs							

TABLE 16

Methods of Assessee Selection

ORGANIZATION TYPE	SUPERVISOR NOMINATION	SELF-NOMINATION	COMBINATION	OTHER
(10) Suppliers		(2) suppliers of educational centers	(1) both	(1) training staff recommendation (1) used tests to pre-screen (5) left choice to client but usually recommends method best suited to client
(3) Consumer Sales	(1) majority of candidates			(1) not specified (1) passed recruiter interview, then self-nominated by accepting assessment (1) let incumbents on target job attend on request for development
(8) Financial	(4)	(2)		(2) not specified
(3) Resource Suppliers	(2)			(1) not specified
(1) Consumer Products	(1)			
(5) Transportation	(5)			
(2) Technical				(1) not specified (1) required for all candidates for target job
(3) Manufacturing				(1) either, felt this is required by law

Table 16 cont.

ORGANIZATION TYPE	SUPERVISOR NOMINATION	SELF-NOMINATION	COMBINATION	OTHER
Manufacturing cont.				(1) either, but usually self-nomination (1) self-nomination, plus personnel office prescreen
(11) Government				
(4) Local		(2)		(2) not specified
(3) State				(1) not specified (1) took applicants from top of civil service register (1) civil service test score, plus two years experience
(3) Federal Agencies		(1) volunteers for project		(1) unclear (1) everyone in certain job classification
(1) Military				(1) all members of training class
(8) Education				
(3) Secondary		(1) self-nomination after system requested applicants		(2) not specified
(5) Postsecondary		(1) chose course, including assessment	(1) included selecting on basis of both	(2) all students in program (1) solicited faculty resumes, then phone interview to choose faculty for assisting paradigm

TABLE 17

The Feedback of Assessment Center Results

ORGANIZATION TYPE	FEEDBACK TO PARTICIPANT				FEEDBACK TO SUPERVISOR				FEEDBACK TO COUNSELOR OR DEVELOPMENT STAFF				OTHER DATA
	Requested:		Given:		Requested:		Given:		Requested:		Given:		
	O*	W*	O	W	O	W	O	W	O	W	O	W	
(10) Suppliers			5	6	1	1	2	2			4	4	(3) depends on client (1) feedback mentioned but form unclear
(3) Consumer Sales			1	1			1	1				1	(1) not specified (1) data to regions but unclear as to whom
(5) Financial			2	2			2	2					(3) not specified
Study norms (N = 21 companies)			✓					✓					
(3) Resource Suppliers			1	1			1	1					(1) not specified but purpose is selection not development
(1) Consumer Products			1				1					1	
(5) Transportation	5			5				5					
(2) Technical													(2) not specified
(3) Manufacturing	1		1	1			1	1					

* O = Oral; Written

TABLE 18

Assessment Follow-up Process

ORGANIZATION TYPE	NO DATA	RECEIVED NO INFORMATION	RECEIVED FEEDBACK	WORKED OUT DEVELOPMENT PLAN ON OWN	WORKED OUT PLAN WITH SUPERVISOR OR ASSESSOR	ONGOING PLAN MONITORED AND UPDATED WITH STAFF PERSON OR SUPERVISOR
(11) Suppliers		(5) left up to client; feedback and development recommended				
			(6)	(1)	(4)	(3)
(3) Consumer Sales			(2)		(1)	(1)
			(1) optional whether it is given	(1) some development of unspecified format		
(5) Financial	(1)	(4)			(1)	
					(1) informal discussion only	
				(1) some development, format unclear		
			(1)*	(1) development less common a goal than is selection		
(3) Resource Suppliers	(1)		(2)		(2)	(2)
(1) Consumer Products			(1)		(1)	
(5) Transportation			(5)			
(2) Technical	(1)		(1)	(1)	optional	intended, but optional
(3) Manufacturing			(3)			

* Norm for survey of 21 companies

TABLE 19

Organizational Use of Assessment Data

ORGANIZATION TYPE	NO DATA	PARTICIPANT FEEDBACK	DEVELOPMENT PLAN	DATA FOR PERSONNEL DECISIONS	KEPT FOR RESEARCH	EVIDENCE OF USE IN RESEARCH
(10) Suppliers		(6) (2) recommended	(4) (2) recommended	(4) (1) users did this although designed for other reasons	(2)	(2) different organization report studies
(3) Consumer Sales		(2) (1) optional depending on request from assessee	(2)	(3)		
(5) Financial		(4)	(2)	(4)	(4)	(1) data-gathering in progress (1) research on pilot in progress by head office (1) report validation study
		(1)*	reported less often	norm	less often	
(3) Resource Suppliers		(2)	(2)	(3)	(2)	(2) validation studies

* Norm. for study of 21 companies

Table 19 cont.

ORGANIZATION TYPE	NO DATA	PARTICIPANT FEEDBACK	DEVELOPMENT PLAN	DATA FOR PERSONNEL DECISIONS	KEPT FOR RESEARCH	EVIDENCE OF USE IN RESEARCH
(1) Consumer Products		(1)	(1)	(1) unintended but supervisor received the assessment data		
(5) Transportation		(5)	(1)	(5)		
(2) Technical		(1)	(1)	(1)		
(3) Manufacturing		(3)		(3)	(1)	validation studies differential validity
(11) Government						
(4) Local	(1)	(2)		(3)	(1)	(1) doing reliability research
(3) State		(2)	(1)	(2)	(1)	
(3) Federal Agencies	(1)	(1)	(1)	(2)	(1)	(1) cited studies in progress
(1) Military		(1)	(1)	(1)		
(8) Education						
(3) Secondary		(3)	(2)	(2)		
(5) Postsecondary	(1)	(3) (1) optional	(2)			

IV. RESEARCH FOUNDATIONS

We have stated throughout this report that the evolution of the assessment center concept into practice has been grounded in a tradition of research. Indeed, many of the top practitioners in the field have been eager to collect data regarding the validity and measurement properties of the assessment center method. This is undoubtedly a key strength underlying current practice, since few other assessment techniques have been subject to such public examination. Nevertheless, as in any research-based endeavor, the results of these studies are not without problems.

We shall attempt to summarize this research with respect to several basic questions to which it has been addressed. It is not possible to detail each study fully, since well over 30 such studies, both published and unpublished, are available . The reader desiring further details would be best served by consulting the original literature or by examining one of the other published reviews of this literature (e.g., Cohen, Moses & Byham, 1977; Huck, in Moses & Byham, 1977; Bray, 1976; Finkle, 1976· MacKinnon, 1975). It is only fair to caution the reader that nearly all of these prior reviews have been authored by advocates of the assessment center concept. This is not to suggest that these reviews demonstrate a heavy-handed bias, but the author's point of view must be taken into consideration in evaluating the interpretations of the data presented. Our own point of view falls somewhat short of advocacy, and is tinged with a degree of skepticism regarding the glowing reports of the assessment center concept as the ultimate solution to assessment problems. With these caveats in mind, let us examine some of the major questions addressed in the research tradition.

Are Assessment Centers Valid?

It is to this issue that the great majority of the research literature has been addressed, primarily as a result of the precedents established in the early research programs on the OSS center and at AT&T. Nowhere is the assessment center literature more complex or more subject to contrasting interpretations. Advocates pronounce the assessment center as the most valid selection method available; skeptics are quick to point to the glaring holes in the evidence. What *is* clear is that there is no easy answer to this question. To even approch a useful answer, we must initially sketch a framework for interpreting the results of this research.

One must recognize that social scientists distinguish several different kinds of validity. The simplest, and least quantitative indicator of validity is *face validity*, i.e., the degree to which the assessment process looks valid to participants or observers. Nearly all published reports on the assessment center concept assert that most assessors and assessees find the process acceptable and valid, particularly in comparison to interviews or psychological tests. The assessment center

Table 17 cont.

ORGANIZATION TYPE	FEEDBACK TO PARTICIPANT				FEEDBACK TO SUPERVISOR				FEEDBACK TO COUNSELOR OR DEVELOPMENT STAFF				OTHER DATA
	Requested:		Given:		Requested:		Given:		Requested:		Given:		
	O*	W*	O	W	O	W	O	W	O	W	O	W	
(11) Government													
(4) Local	1		1	2				2					(2) piloting for selection; no information on feedback
(3) State	1			2				2				1	
(3) Federal Agencies	1	1											(1) unclear (1) pilot no feedback
(1) Military			1								1		
(8) Education													
(3) Secondary			2	1			1	2					(1) format not specified through package for development
(5) Postsecondary	1		3					1			1		

* Oral; Written

Table 18 cont.

ORGANIZATION TYPE	NO DATA	RECEIVED NO INFORMATION	RECEIVED FEEDBACK	WORKED OUT DEVELOPMENT PLAN ON OWN	WORKED OUT PLAN WITH SUPERVISOR OR ASSESSOR	ONGOING PLAN MONITORED AND UPDATED WITH STAFF PERSON OR SUPERVISOR
(11) Government						
(4) Local	(2)		(2)			
(3) State		(1)	(2)	(1)		
(3) Federal Agencies	(2)		(1)			(1)
(1) Military			(1)			(1)
(8) Education						
(3) Secondary			(3)	(1) (1) development plan prescribed for each candidate but not face-to-face or followed up		
(5) Postsecondary		(1)	(4) (1) optional	(3)	(2)	

process clearly passes this test with flying colors in most instances, and little more will be said about it.

A related measure of validity is content validity, i.e., the degree to which the items or procedures in an assessment technique sample the performance content of the criterion job or situation towards which the assessment is directed. Given the proper design of the standardized performance situations as a function of prior job analyses, the assessment center process is, in effect, constituted from work samples from the job. In this respect, the assessment center is probably more content valid than many other traditional assessment procedures. However, content validity is not easily quantified and typically must be based on some form of human judgment. Given the trend in recent EEO guidelines and judicial decisions regarding employment testing, it is likely that judgments of content validity will need to be quantified in some manner. At a minimum, assessment center administrators will need to produce more details on the procedures followed in the job analysis upon which the center is based. Such procedures, and their results, are rarely published at present. In addition, measurable criteria will have to be developed which can inform the interested observer of the extent to which the perfomance situations in the center adequately sample the universe of situations found in the criterion job or jobs. Unfortunately, from our point of view, this may occur at the expense of more informative validation research.

A third type of validity relevant to assessment center research is *construct validity*; i.e., the degree to which an assessment technique validly measures the constructs (i.e., traits, characteristics, skills, abilities) which it is designed to measure. The concept of construct validity is both tricky and difficult to operationalize into measurement. The reason for this is that we have very few, if any, ultimate (or objective) measures of particular constructs associated with human behavior to serve as criteria for validating a new measure of the construct. Campbell and Fiske's (1959) concept of the multitrait-multimethod matrix, upon which the assessment center technology is fundamentally based, can be, and has been, used for construct validation. The basic notion here is that two measures designed to measure the same construct should be more intercorrelated than two measures designed to measure two different constructs. In effect, construct validity is an attempt to evaluate what an assessment technique "really" measures.

There have been relatively fewer attempts to assess the construct validity of assessment centers. One reason for this is the fact that most research has focused more on the overall competency judgment as a predictor of some criterion of job success than on the correlates of specific behavioral dimension ratings. In addition, practitioners tend to avoid describing behavioral dimensions as underlying constructs of the individual. Most subscribe to the notion that observable behavior in the center is used to predict future behavior on the job; underlying constructs are viewed as unnecessary and problematic. One could argue that this viewpoint represents something of a semantic game, particularly since observable behavior must be rated on a dimension which represents some attributed quality of the individual. Avoidance of construct validation is simply not justifiable.

The construct validation research which has been undertaken most often takes the form of statistical clustering (via a method called factor analysis) of assessor ratings of the behavioral dimensions. Other researchers have attempted to examine specific performance situations such as the in-basket or leaderless group discussion in order to assess the characteristics of the individual which seem to be truly measured. Others have attempted to assess the differential contribution of assessment center methods and results, in conjunction with more traditional measures, in predicting job success. Few researchers have employed the strategy of correlating the dimensional ratings of an assessee with ratings of his or her actual performance on the job in similar situations. This is clearly a weakness in the current research base. We shall return to a review of these studies in the latter part of this section addressed to the question of what an assessment center truly measures.

The final type of validity, towards which most assessment center research has been directed, is criterion-related validity, i.e., the degree to which the assessment technique accurately predicts some criterion of interest, either in the present (concurrent validity), or at some time in the future (predictive validity). Predictive validity is, of course, the "bottom line" of validation research, and is the "common sense" interpretation of the concept of validity, particularly in work settings. Nevertheless, the conduct of predictive validation research is more complex than it might appear. The reader must take into account several critical aspects of this type of research.

First, one must understand the nature of the variables from the assessment center that are used to establish its predictive power. Very few published studies utilize the ratings on the behavioral dimensions as the primary predictors of the criterion measure. In fact, many studies report only moderate correlations between the dimensional ratings and the criterion. Rather, the great majority of the predictive validity studies rely primarily on the final, overall competency judgment, however quantified, as the primary predictor of the criterion. These final ratings are typically made on 2-point to 5-point scales, and represent some form of reduction, or summary, of the multiple dimension ratings. Thus, as we have noted before, the manner in which this final rating is obtained is fundamental to the predictive power of the assessment center method. In yet other studies, the authors simply compare the job performance of a group promoted following an assessment center experience with a group promoted by more traditional methods. These studies are far less informative, in general, to the validity of assessment due to the absence of specific measures of predictor-criterion relationships.

Second, one must closely examine the criterion measure which the assessment center attempts to predict. Many claims for the validity of the assessment center are typically assertions that the method can predict "success in management." This criterion has typically been represented by such objective criteria as growth in salary, growth in managerial responsibility, rate of promotion, or management level achieved over time; or by such subjective criteria as supervisory performance appraisal ratings or peer ratings. These are readily available criteria, but one must note that, for the most part, they are not equivalent to criteria measuring actual performance on the job. Very few studies attempt to validate the assessment center process by predicting actual job perfomance. Bray and Campbell (1968) and Clingenpeel (1979) conducted two such studies. Unfortunately, the former study collapsed observations of actual job performance to one binary measure of "meeting performance standards" rather than measuring multiple behaviors on the job. The latter study found very small correlations between assessment center ratings and some 40 different performance criteria. The attempts to use supervisory ratings of performance as criteria have also proved troublesome since these ratings tend to be unreliable and inconsistent with other ratings by outside observers (see Bray and Campbell, 1968, or Thomson, 1970, for examples of this phenomenon).

In short, the predictive validity of the assessment center has been typically based on the more remote outcomes of performance, not on performance per se. Other writers, including Hinrichs (1978) and Klimoski and Strickland (1977), have reached similar conclusions. The latter authors, upon examination of the research base, concluded:

> While these popular criterion measures are intercorrelated and appear to measure advancement or success within an organization, it is not clear that they reflect competence, effectiveness or superior performance. . . We want predictive validity studies of performance (1977:336, 359).

Third, the design of predictive validity research can suffer from various methodological flaws, of which two appear most critical. Nearly every published study, save three (Bray, Campbell & Grant, 1974; Bray & Campbell, 1968; Hinrichs, 1978) has suffered from what has been called the "criterion contamination problem." Simply stated, most research has been conducted on operational centers where the results of the center have been used by the organization to make promotion and selection decisions. This, of course, can lead to an artificial elevation in the predictive validity of the center. To overcome this problem, some researchers have compared the results of assessment center selection versus other methods used prior to center implementation (thereby comparing assessed versus nonassessed groups), or have examined the number of promotions of the assessee beyond the management level for which the assessee was originally assessed. Neither strategy is without problems; the former may compare two distinctly different groups while the latter may suffer from biases introduced by supervisory actions towards the assessee based on his or her performance in the center (e.g., with respect to job placement or degree of responsibility assigned to the assessee on the job).

The utilization of assessment results by the organization can also lead to what has been called "the restriction of range problem." This simply means, for example, that if assessees are promoted based on the results of the center, there is likely to be less variance in their advancement rates (i.e., the criterion) compared to their varied performance in the center. Thus, validity results can be spuriously affected, since there is little to predict. Nevertheless, in the great majority of the research studies, it is difficult to judge the extent to which these two types of design problems affect the results.

Fourth, one must understand the difficulties inherent in comparing the results of predictive validity research since these studies employ varied predictors, criterion measures, and statistical tests to relate predictors and criteria. The easiest statistic to grasp is the simple correlation coefficient of the the predictor (most commonly the 2-point to 5-point overall rating) with one or more criterion measures. Such correlations can range between -1.0 to +1.0. A negative correlation means that the predictor and criterion are inversely related, i.e., the higher one's score is on the predictor, the lower one's score is on the criterion. A zero correlation means the two measures are unrelated, and a positive correlation means that high scores on the predictor are related to high scores on the criterion. There are no fixed standards to establish how large (i.e., greater than zero in either a positive or negative direction) such correlations should be before they are considered acceptable support of validity. However, social scientists would probably agree that correlations in the range of .20 to .35 define low-to-moderate validity relationships. Many standard personality and aptitude tests tend to correlate at this level with performance criteria. Correlations in the vicinity of .40 to .60 are generally considered moderate to high, and are usually viewed as sufficient evidence of validity. Correlations above .60 are quite high; in fact, when the correlations obtained in research are much higher than .60, social scientists are often suspicious that predictors and criteria are spuriously related (e.g., the two measures are really identical or a third variable causes their apparent relationship). One final note on correlation coefficients: one can quickly establish how much of the variation in the criterion measures can be accounted for, or "explained," by the predictor. Simply square the correlation coefficient. For example, a

210

validity coefficient of .50 suggests that 25 percent of the variation in the criterion can be explained by variation in the predictor.

Unfortunately, not all studies report correlation coefficients. Many present the results in contingency table form (see, for example, the results of the Management Progress Study in Section II of this report). In this format, the number, or percentage, of assessees receiving overall ratings at each point of the rating scale is tabulated against the number or percentage who fall at certain points on the criterion measure scale. For example, the percentage of assessees receiving high assessment ratings and who receive promotions can be compared with the percentage of assessees who receive low assessment ratings and are not promoted (of course, the inverse relationships can also be examined, e.g., high ratings versus no promotion). Comparison of these percentage figures is often compelling evidence for validity to many readers of the literature. Yet it is difficult to translate these results into validity coefficients since the two different measures are based on different assumptions regarding the numerical properties of measurement scales. We would simply caution the reader not to be misled by the presentation of percentage differences; the actual amount of variation explained may be somewhat lower. Where possible, ask for correlation coefficients.

Fifth, and finally, one must examine the population of subjects upon which the research is conducted in order to assess to what degree the results are generalizable. At this point in the evolution of the research foundations of the assessment center concept, the great majority of published validation research comes from a small number of programs in larger corporations (especially AT&T, IBM, SOHIO and Sears). It may be somewhat hazardous to generalize the results of this research to other organizations; furthermore, the courts may not find previous validity research to be applicable to other applications. This fact merely suggests the critical need to expand the research base as much as possible. As we noted in our survey of practice, such studies are in progress at a number of companies, and it is our hope that these results will be published in the near future.

Keeping in mind the different types of validity, and the five points associated with the design of predictive validation, the reader should now be in a better position to interpret the results of the many research programs addressed to the issue of assessment center validity. Let's review these results briefly.

The AT&T Research

The conduct of the Management Progress Study at AT&T led to heightened interest in the assessment center method as an operational tool for selection, particularly of first-level supervisors with potential for higher management. By 1958, Michigan Bell had modified the research procedures of the Management Progress Study (specifically excluding personality tests and projective measures requiring professional psychological interpretation) to assess the managerial abilities of craftsmen. This center has served as the prototype for many other operational programs both within and outside of AT&T, and has resulted in the publication of several evaluation studies. Despite the fact that the majority of this research suffers from criterion contamination, several interesting patterns appear in the results.

An early study by Michigan Bell (1962, and discussed in Bray, 1964) of its center compared the first 40 men assessed in the center, and promoted to first level supervision, with the last 40 men promoted prior to implementation of the center. The company found that approximately two-thirds of the assessed group were rated on their jobs by their managers to have displayed "better than satisfactory" performance and rated to possess the potential for higher management jobs. Only one-third of the nonassessed group received such managerial ratings. This study strongly suggested that the assessment center was more predictive than other selection methods; nevertheless, the managerial ratings could have been biased by knowledge of assessment results, and the two groups may have been unevenly matched.

Mather (1964) compared assessees receiving high and low overall ratings (from New England Bell's assessment center) who were subsequently promoted to first or second level supervisory positions. Managerial ratings of performance served as the criterion. Seventy-six percent of the highly rated assessees were evaluated by their managers as displaying "high-quality job performance"; forty percent of the low-rated assessees received such a managerial rating.

In 1965, AT&T conducted a study of the consequences of centers in four other Bell companies (summarized in Campbell & Bray, 1967). A total of 506 individuals, all of whom had received promotions, were examined. This sample was composed of five distinct groups. Three groups, (n = 223) had been assessed and had received overall ratings of acceptable, questionable, or not acceptable for selection. Two groups (n = 283) were not assessed, but had received promotions either prior to, or after, the implementation of a center. The criteria included their last supervisory performance appraisal by their middle-management supervisor. The welter of comparisons presented can be reduced to a few basic conclusions. First, a comparison of assessed versus nonassessed groups showed that assessment, per se, led to slightly higher ratings of performance. Within the assessed groups, significant differences were found in the managerial ratings of performance between those assessed as "acceptable" and "not acceptable." However, by far the greatest difference was in the managerial ratings of potential for advancement. Approximately one-half of those assessed as "acceptable" were rated as having high potential for advancement versus only one-quarter of those assessed as "not acceptable."

A particularly unique study was reported by Bray and Campbell (1968). Seventy-eight candidates for sales positions participated in a two-day center designed on the basis of an extensive analysis of sales jobs. The assessment results were not shared with management. Predictors were 20 dimension ratings and the overall rating (on a 4-point scale ranging from *more than acceptable* to *unacceptable*). A special field-review team was created to observe the assessee's actual performance on the job. These observations were used to construct a binary criterion measure of "performance meeting standards, or failing to meet standards." The strongest result was the correlation of the overall rating and this criterion measure of performance (.51). The study also pointed to the difficulties associated with the use of supervisory ratings of performance as the criterion. The assessment center ratings did not accurately predict these supervisory ratings. In effect, the closer examination of the performance of the assessee by the special review team increased the validity of the assessment predictions, even though this criterion measure represents a rather considerable reduction of performance variations to a 2-point scale.

Moses (1972) presented an impressive summary of data based on an enormous sample of individuals assessed in operational centers by Bell companies throughout the 1960's (n = 5943). The primary predictor was the overall rating received by the assessee (on a 4-point scale). The criterion selected was advancement beyond the level for which the assessee had been originally assessed. The assumption here was that such a criterion would be less contaminated by the bias in supervisory ratings of performance or by the bias introduced by supervisory knowledge of the assessment results for the first promotion. Of those individuals rated as "more than acceptable" in a center, 40.5 percent had received two or more promotions versus 21.5 percent of those rated as "acceptable"; 11.5 percent of those rated as "questionable"; and 4.2 percent of those rated as "not acceptable". The correlation of this overall rating with this criterion of management progress was moderately high (.44).

Moses and Boehm (1975) conducted a similar study based on a sample of 4,846 women assessed between 1963 and 1971, and a very similar pattern of results emerged. The overall assessment rating correlated .37 with management level, even though the majority of this sample had been assessed more recently (the average time of assessment was 1969), thereby decreasing the possibility of 2 or more promotions. In another study, Huck and Bray (1976) found that the overall ratings of black and white female assessees correlated moderately well with managerial ratings of the assessee's overall effectiveness on the job (.41 and .35 for the two groups, respectively) and quite well with managerial ratings of the assessees' future potential for advancement (.59 and .54, respectively).

Taken as a whole, this series of studies within the Bell System constitute a fairly impressive confirmation of the predictive validity of the assessment center method, particularly when the predictor is the overall rating from the center and the criterion is management level achieved over time, or global managerial ratings of the assessee's job performance.

The IBM Research

Wollowick and McNamara (1969) followed up a group of 94 assessees from lower- and mid-level management some three years later. The predictors included 14 different paper-and-pencil test scores, 6 assessor ratings of performance in specific assessment situations, 12 assessor ratings of behavioral dimensions, and the overall rating. The criterion employed was a measure of increase in managerial responsibility, defined through a job analysis to include such factors as number of subordinates, job complexity, and fiscal responsibility. The multitude of correlations showed that the overall rating from the center tended to be the best predictor of the criterion (.37), though some of the other predictors reached nearly equivalent validity. The authors also discovered that a statistical weighting of the predictors proved vastly superior to the overall rating in correlating with the criterion (.62). Thus, this study suggested that the clinical aggregation of the assessee's performance in the center by the assessors (as represented in the overall rating) was not as valid as an empirical combination. Unfortunately, this result has not been consistently replicated, and much more research needs to be conducted on this issue (see Bray & Campbell, 1968; Moses, 1972; Mitchel, 1975).

Dodd (1971) summarized some 11 studies conducted at IBM involving a total sample of 479 assessees. He found a median correlation of .35 between overall ratings from the centers and such criteria as position level and salary or changes in position level and salary. One study employed the unique criterion of demotion rate among assessees following entry into a new job. Only 4 percent of highly rated assessees were subsequently demoted versus 20 percent of low-rated assessees.

Hinrichs (1969, 1978) followed a group of 47 assessees over an 8-year period. Predictors included 12 dimensional ratings and an overall rating of potential from the center (as well as a variety of paper-and-pencil test scores); in addition, he obtained similar overall ratings of potential from two managers who simply reviewed the assessee's personnel folder with no knowledge of the assessment results. The primary criterion was position level at year 1 and at year 8. The correlation of the overall rating and position was .26 (for year 1) and .46 (for year 8). Interestingly enough, the ratings based on simple review of folders correlated equally well, if not slightly better, with management level than did the overall assessment ratings (.32 and .55, respectively). Hinrichs also found significant predictive validities for several of the specific dimensional ratings (range = .34 to .69), as well as for certain scales from the objective tests.

Hinrichs' conclusions are intriguing. While acknowledging the growth in the predictive validity of the assessment center over time, he suggests that the equally good results obtained by the managers reviewing personnel folders raises the possibility that...

> ...both methods are essentially policy-capturing devices or proxy measures of what happens under natural conditions in the organization. Perhaps the managers who were assessment observers and the managers evaluating the personnel files were factoring into their judgments personal knowledge of the promotion criteria and standards used in the organization (1978:600).

The bottom line here is that use of assessment centers to predict only the criterion of advancement may be an overly expensive and complex methodology. However, prediction of actual performance using assessment center data may be justifiable, even though this study, like so many others, did not employ performance criteria for validation.

Finally, Kraut and Scott (1972) examined the career progress over two years of nearly 1100 employees in sales, service, and administrative functions following their participation in a center. The primary predictor was the overall rating of potential from the center. The primary criterion was promotion to higher levels beyond the level for which the assessee had been originally evaluated. In all three functions, a greater percentage of the highly rated assessees received further promotions than the lower rated assessees, though this difference was moderated by the general rate of promotion within each function. The results of this study are quite supportive of the results reported by Moses (1972) at AT&T. The authors also found, for the sales function at least, that demotions from first-level management were related to the assessment rating, though this result was only of marginal statistical significance.

The SOHIO Research

Carleton (1970) followed up a sample of 122 assessees over a 2-1/2 to 5-year period following participation in SOHIO's entry-level management assessment center. Finley (1970) used the same program to follow up an additional sample of 109 assessees over a shorter period. Both studies employed 13 different predictors, including ratings from different assessment exercises as well as an overall rating based on this information and further scores from paper-and-pencil tests. The criteria included management progress (Carleton's study only), 12 managerial ratings of performance, and 1 managerial rating of future potential. The results were quite similar for both studies. The overall rating from the assessor team tended to be the best predictor of the criteria, though the correlation coefficients varied across the criteria. Specifically, the overall rating was moderately correlated with the 12 managerial ratings of performance (average = .25 and .32 for the Carleton and Finley studies, respectively); strongly correlated with progress in the Carleton study (.51); and highly correlated with managerial ratings of future potential in both studies (.65 and .63, respectively).

Thomson (1970) presented the results of a complex multitrait-multimethod analysis of one SOHIO program. Seventy-one assessees were evaluated in a center by both psychologists and managerial assessors. Independent ratings by these 2 groups of assessors were made on 13 predictor dimensions including an overall rating of potential. The criteria were supervisory ratings of the assessees on the same 13 dimensions from 6 months to 2 years following participation in the center. The ratings made at the assessment center by the psychologists and the managers were highly correlated (median = .85). However, the predictive validities for both groups to the criterion supervisory ratings were substantially lower (median = .42) and .38 for psychologists' and managers' predictions, respectively). Assessment predictions of potential were much more highly correlated (.64 for both groups) with supervisory judgments of potential. Thomson discovered from further analysis that the supervisors were far more lenient in their ratings, used a more restricted range of scale values (on a 9-point scale), and were far less reliable. In addition, the supervisors were not able to adequately discriminate ratings on the various dimensions. This study serves to reaffirm the difficulties associated with the use of supervisory ratings of performance as criteria for validating an assessment center, even though the predictive validities obtained were moderately high.

Mitchel (1975) gathered data on some 254 assessees evaluated in centers over a 6-year period. Thirty-three different predictors were examined, including paper-and-pencil test scores, peer ratings on several dimensions, and assessor ratings on 8 dimensions (plus an overall rating of potential). The criterion measure was a complex measure of salary growth calculated at 1, 3, and 5 years from assessment. Tests were found to correlate quite poorly with this criterion. Neither peer ratings nor assessor ratings were, on the average, highly correlated with this criterion (.17 and .24, respectively). Mitchel did find a trend for increased predictive validity as one moved further in time from the assessment. Interestingly enough, the validities for the earliest assessees assessed in 1966 to 1968 were much higher (average = .33) than for later assessees assessed in 1969 to 1972 (average = .14). No interpretation is offered for this phenomenon, but one might speculate that changes in assessor training or other procedures had a great impact on the validities of different centers. This simply highlights the need to update the validation of a center continually, especially when any changes in procedures are implemented.

The Sears Research

Sears and Roebuck has employed various tests and procedures for entry-level management selection since the 1930's when

L.L. Thurstone was retained to develop a paper-and-pencil executive assessment battery. In the early 1960's Sears turned to the use of assessment centers as one component of this selection process. Under the direction of V. Jon Bentz, this application has resulted in a longitudinal study of assessment and management behavior, from which data will be derived for years to come. This effort demonstrates one of the better attempts we have seen to carefully design the overall research process, and to rigorously test every assessment technique. Revisions in the assessment process continue to occur as a result of the research. Much of the data from this research has yet to be published, though the Sears researchers are presently preparing analyses for publication. We look forward to their publication.

Nevertheless, Bentz (1971) and Marquardt (1976) have presented the results of some initial validity studies. In one initial validity study based on a small sample (n = 51) of executives, Bentz employed a variety of predictors from several personality tests, an in-basket, and two group problem-solving simulations. Criteria included ratings made by Personnel Directors of effectiveness and potential; self-ratings; ratings by one's immediate superior on potential, creativity, and style; and two objective indices of job mobility and salary progress. Results were presented for each type of assessment technique.

The in-baskets were scored on a variety of dimensions. Correlations between these dimensions and the various criteria typically ranged between .28 and .37, though correlations with managerial global ratings of effectiveness were much higher (.41 to .50). Multiple correlations of these dimensions of in-basket performance with criteria were also quite high (.35 to .50). A virtually identical pattern of results was found in analyzing peer and observer ratings of assessee performance in the group problem-solving simulations. One intriguing result of this study was that many of the multiple correlations of personality test scores with criteria produced equally high, if not higher, validity coefficients (.40 to .60)! The results of this initial study encouraged Bentz and his colleagues to proceed with the larger longitudinal design presently under evaluation.

Bentz (1971) also presented some initial concurrent validity findings for one sample of 100 management trainees in this longitudinal study. The predictors were very similar, but criteria included training performance evaluations at 6 months and 1 year, as well as performance ratings on 52 different elements by store managers to whom trainees were assigned. Another welter of validity coefficients are available; however, we shall focus primarily on the job performance ratings. These ratings were reduced to 6 common factors. Bentz found that scores from the mental ability tests were generally unrelated or negatively related to these performance rating factors. Various dimensions of the in-basket correlated moderately low with these factors (.23 to .30), as did various predictors from the group problem-solving simulations (.24 to .30).

It should be noted that Bentz has not examined the validity of any overall rating from the assessment center, but rather has chosen to focus on more specific predictors from each of the assessment techniques. While the results have encouraged the developement of further research, they are not as consistently strong as the results emerging from the validity studies of overall ratings. In addition, these studies provide strong support for the validity of some paper-and-pencil instruments. A real value of this data base is the degree of rigor applied to its design and development. We suspect that many intriguing results will continue to emerge from this program in the next few years.

Other Validity Studies

Several other studies have been published, or at least presented to symposia, which bear upon the issue of the validity of assessment centers employed in a variety of other organizations. Let us briefly review these studies.

Dicken and Black (1965) assessed managers in both an insurance and a manufacturing operation. They found an overall rating correlated at .58 with salary growth following assessment. Albrecht, Glaser and Marks (1964) used a series of paper-and-pencil measures and interviews to assess a very small group of marketing managers (n = 31). Criteria included managerial ratings of performance. Assessor (all consultants) rankings of assessees on the criterion dimensions (sales, forecasting, and interpersonal ability, as well as overall performance) correlated with managerial ratings at .27 to .46. This study suggested that a clinical summary of paper-and-pencil tests can predict managerial ratings of performance as well as simulations can.

Glaser, Schwarz and Flanagan (1958) studied civilian supervisors at military depots (n = 80). Predictors include assessor ratings from individual and group interviews, a group discussion, a role-play situation, and two paper-and-pencil tests. Criteria were managerial ratings of performance as well as a managerial record of on-the-job performance. Individual exercises correlated only from .12 to .17 with these criteria, though various multiple correlations raised these coefficients to a moderate level of validity (.30 to .33).

In the late 60's the Peace Corps became involved in the assessment of volunteers for overseas assignments. Many of these efforts emphasized paper-and-pencil tests or board interviews (see Harris, 1972) as predictors. However, Gordon (1967) attempted to compare the predictive validities of paper-and-pencil tests, situational simulations, and work samples (language acquisition training) in one study where the criterion measure was selection of the volunteers for overseas assignment by the review board following training. Gordon found that all sets of predictors were of equal validity (.37 to .41) in predicting selection for a sample of 178 volunteers, though each method was somewhat independent of the others.

Bullard (1969) presented one of the few studies comparing selection based on an assessment center with normal procedures, in which the center predictions of job success were less valid than prior selection procedures. He compared the job performance ratings of 37 individuals promoted after assessment center participation with those of 27 individuals promoted prior to center implementation. Correlations between assessment ratings and supervisory ratings of job performance ranged from -.09 to .26 and were not significant. Unfortunately, these groups were not well matched (e.g., only marginal performers appeared to attend the center; superior performers were promoted without assessment).

Jaffee, et al. (1970) conducted a similar comparison of assessed versus nonassessed first-line supervisors for a small application (n = 26). The authors found that subordinates described the supervisors selected through an assessment center much more positively than supervisors selected through other means. Meyer (1972) found that 16 percent of the first-line supervisors in his sample at General Electric were subsequently demoted when no assessment was utilized in contrast to no demotions in a group promoted after participation in a center.

Ginsburg and Silverman (1972) compared the results of a one-day assessment center for 37 hospital administrators with their present job performance ratings. The median correlation of the assessment center dimension ratings and the job performance ratings was reported to be .34. Thornton (1971) reviewed the available data from a number of smaller companies employing the assessment center method. While there were not a great amount of validity data available, he did find 3 studies where the median correlations of predictors and criteria ranged from .20 to .30. In certain cases, these validities were considerably higher (.50 to .60).

McConnell and Parker (1972) presented the results of a multi-organizational application of an assessment center designed by the American Manangement Associations. These authors found overall ratings from the assessment center to correlate at .57 with job performance ratings of 70 first-line supervisors in 4 different organizations. Worbois (1975) reported a study of the AMA center as it was applied to one organization. Forty-eight first-level supervisors were assessed in a center on 12 separate dimensions and received overall ratings on potential for successful supervision and potential for advancement. Criteria included managerial ratings on the 12 dimensions and an overall rating of performance effectiveness. The dimensional ratings correlated in the vicinity of .35 to .45, while the overall ratings from the center were correlated .45 to .47 with the various overall managerial ratings. These two studies demonstrated strong support for the validity of the AMA program. A more recent analysis by Parker (1978) of a somewhat larger sample of participants in the AMA program (n = 280) used assessment center ratings to predict job performance ratings employed by various organizations. The 12 dimensional ratings were found to be moderately correlated with these performance ratings (.26 to .33), while an overall rating of management ability from the center correlated at .36 with ratings of performance. The somewhat less strong results of this more recent analysis are probably due to the variance in performance evaluation scales employed across the several organizations in the sample. Parker was forced to reduce these variations to a single 3-point scale for maximum comparability. Unfortunately, this produced the unwanted side effect of reducing variance in the criterion and thereby lowering the validity coefficients. Nevertheless, the AMA program has produced a considerable data base (over 3,000 assessees in the past 10 years) which holds great potential for future research analyses.

The Public Service Commission of Canada has been actively involved in the use of assessment centers since the early 1970's. Slivinski, Grant, Bourgeois, and their colleagues have been collecting data on several applications designed over several years (basic materials can be obtained from the authors--see especially Slivinski et al. 1977). Two longitudinal studies on samples of 45 and 72 assessees attending a first-level management assessment center were presented in Slivinski, et al. (1978). Predictors included assessor ratings of 12 dimensions and an overall rating of potential for success at first-level management. Criterion measures were obtained at 1-1/2, 3, and 4-1/2 years from assessment for the first sample and at 2 and 3-1/2 years for the second sample. Criteria included salary, position level, a departmental appraisal, and supervisor's overall performance evaluation. Various combinations of these criteria were devised, and supervisors used specially designed behaviorally-anchored rating scales to rate the assessee on the same dimensions assessed in the assessment center. For the first sample, the overall rating from the center correlated quite highly with the four criteria, and their various combinations (range = .42 to .50), as measured at the different time points. The correlations of the dimensional ratings of assessors with those made by supervisors showed much greater variability. However, one compelling pattern was apparent: the validities were much higher at the 3-year follow-up than at the 1-1/2 or 4-1/2 year follow-ups. This was the point at which actual job mobility had begun to occur and assessees were beginning to reach the jobs for which the assessments had been targeted. Nevertheless, this result suggests that researchers conducting longitudinal evaluations of validity may find validity coefficients varying considerably over time.

Interestingly enough, the same analyses on the second sample produced a far less compelling result. Only one criterion (salary increase at 3-1/2 years) was correlated above .40 with overall ratings from the center; all other correlations ranged between .05 to .23. The predictive validities of the dimensional ratings were equally low (the majority were in the vicinity of .10 to .30). These results once again suggest that the validity of an assessment center may not be able to be established on a permanent basis. The authors themselves pointed to differences in assessee group composition and assessment procedures which may have produced these differences. Needless to say, this research group will surely contribute valuable analyses in the future. (For example, Slivinski et al. (1979) used a sample of 278 assessees to show that a combination of paper-and-pencil tests, biographical data and simulations were better predictors of upward mobility than any one technique by itself.) We look forward to further analyses of this data base, one of the few available in government settings.

Finally, we were able to identify four other recent unpublished studies of validity in four diverse settings. Dyer and Hilligloss (1977) used a 2-1/2-day assessment center to predict field leadership performance of Army Infantry officers. Predictors included objective test scores, assessor ratings on each of the exercises developed, and a variety of peer and self-ratings. Criteria were performance ratings on 10 dimensions completed by supervisors, peers, and subordinates of the assessee at 6 months and 18 months following assessment. The researchers were generally disappointed in the results due to the fact that assessor ratings tended to be very poorly correlated with the criterion ratings in general. They also found that for different subgroups of assessees different patterns in the predictiveness of assessor ratings were displayed, suggesting once again a certain instability in the predictor-criterion relationships across assessee groups.

Silzer (1978) designed an assessment center to measure competencies for pharmacists. An average validity coefficient of .17 was found between exercise ratings and performance ratings (by two peers of the assessee). Dimensional ratings were slightly more valid (average = .25). However, the performance ratings utilized were highly unreliable. This research, however, is only at the very initial stages. Strickland (1979) has conducted an intriguing thesis study in which he has attempted to measure the impact of center implementation on a large insurance firm. He discovered validity coefficients between overall assessment rating and such criteria as promotion or salary increase to range from .15 to .37, with the highest correlations apparent between the overall rating and managerial ratings of future potential. No significant relationship was found between assessment ratings and managerial ratings of performance.

Clingenpeel (1979) reported similarly weak results from a recent study at General Motors on the validity of a foreman selection process utilizing a one-day assessment center. The sample included nearly 800 assessees from 23 plants (from a larger sample of approximately 4300 assessees across 80 plants). The primary predictors included assessor ratings on 8 dimensions and an overall rating, though the total number of predictors was 340 since each performance situation included observation forms with distinct variables. Some 40 different criteria were employed, including performance appraisal ratings, personnel file data (promotions, salary progress, etc.), and special ratings by the assessee's immediate supervisor collected for the purposes of this study. The correlations of predictors and criteria were uniformly quite low. The best correlations were obtained between assessment ratings and supervisory ratings of the assessee's future potential (.20 to .25). Using the strongest predictors in combination did produce multiple correlations in the range of .23 to .34. The author did find some evidence for increase in validity the further in time the criteria were collected from assessment (supporting the results of Mitchel [1975] and Slivinski [1978]), and speculated that collection of criteria only 1 year or less from assessment seriously limited the validity coefficients. Nevertheless, these results have led to a wholesale revision of the program and a new revalidation effort.

Conclusions of Validity Research

What can be said at this point of the validity of assessment centers? *A reasonable answer is that there is a fairly stable pattern of moderately good predictive validity over a range of studies; however, the nature of this validity must be closely examined due to the variation in the research procedures and results*. Several points are worth emphasizing.

First, there is not conclusive evidence that certain paper-and-pencil measures are less valid than assessment center predictions. Several studies show comparable validity coefficients, and some show higher validity for tests. We suspect that this issue is a red herring. The more important issue is assessing the differential contributions of these different techniques to validity. This viewpoint is shared by some recent writers (e.g., Slivinski, et al., 1979). Much more research needs to be directed to this problem.

Second, the best predictor from the assessment center is typically the overall rating as opposed to the ratings of the specific behavioral dimensions, or ratings based on specific performance situations. Since this overall rating is usually not a strictly-defined or weighted summary of the dimensional ratings, its meaning as a predictor is open to some question. Comparisons of the validity of these overall ratings when obtained by judgment versus when obtained mathematically are not conclusive. Therefore, the decision rules used to make these ovverall ratings must become more explicit; otherwise we will be left with a certain lack of clarity as to what this overall rating really measures.

Third, a variety of criteria have been employed to establish validity. As Cohen, Moses and Byham (1977) have previously suggested, the assessment center method tends to best predict global managerial ratings of future potential (correlations often range as high as .65). Objective measures of management progress, such as salary growth or level attained, are predicted fairly well (correlations typically range in the low .30's to the high .40's). Measures of actual job performance tend to be predicted least well (correlations typically range between .20 to .40). *In short, assessment centers tend to predict predictions of future potential better than actual job performance or objective measures of management progress or "success*." As we have stated previously, such studies of performance must be conducted in order to strengthen the research base--particularly given the great difficulties apparent in using managerial *ratings* of performance.

Finally, there would appear to be a greater variation in the results of more recent research, and certainly more studies finding lower validity coefficients. The reasons for this phenomenon must be closely examined. There is, at least, a hint that the validity of an assessment center is not stable over time (e.g., Slivinski et al., 1978). Furthermore, changes in procedures can produce very different levels in validity (e.g., Mitchel, 1975). This will require far more research before the true predictive validity of the assessment center concept can be adequately assessed.

Are Assessment Centers Reliable?

The issue of measurement reliability is a precursor to the establishment of validity. The basic question here is whether the assessments produced show some stability or consistency. This can be measured in three ways. *Inter-rater reliability* evaluates the degree of agreement among two or more raters observing (and rating) behavior in a performance situation. Inter-rater reliability can also be calculated for summary ratings of dimensions and for overall ratings, but only when these ratings are made independently by assessors (i.e., prior to discussion). *Test-retest reliability* evaluates the degree of stability in assessments over time (i.e., repeated measurements over short periods of time, typically). *Internal consistency*, as applied to assessment center research, is a variant of reliability which evaluates the degree to which a dimension measured in one performance situation is highly correlated with the same dimension measured in other situations. In effect, this type of reliability tests for the extent to which a dimension is consistently measured by alternative techniques in the center. (The knowledgeable reader may recognize this as identical to the concepts of construct validity or convergent validity. In fact, the use of the internal consistency concept is somewhat inappropriate in this context since it is typically applied to the interrelatedness of items within one test. Nevertheless, if one considers the assessment center as a global test attempting to measure specific variables through multiple techniques, then a measure of the degree to which these variables are measured consistently across techniques can be viewed as a form of reliability.)

All of these types of reliability are typically measured through correlation coefficients. In contrast to validity coefficients, reliability coefficients are commonly expected to be much higher. Psychometricians usually expect reliability estimates to fall, at a minimum, above .70; and many prefer such estimates to range above .80 in order to support the assumption of reliability. Correlations below .70 usually raise some concern for redesign of the instruments employed. Table 20 presents a summary of the reliability research on assessment centers. Several points warrant discussion.

First, reliability estimates can be calculated for ratings made in specific performance situations, for dimensional ratings, or for overall ratings, though the second alternative is most often selected. There would appear to be a slight tendency for these dimensional reliabilites to be somewhat lower, on an average, than the reliability of the overall rating. Nearly every study reports some particularly weak dimensions. This, of course, necessitates redesign of the dimensions, the performance situation, the assessor rating format, or even assessor training. Without such redesign, validity is seriously constrained.

Second, inter-rater reliability is the most common form of evaluation. These correlations are usually of minimum acceptability (.70 or higher) though a degree of variation is apparent both within, and across, studies. The one true test-retest evaluation (Moses, 1973) found reliability estimates that ranged from .49 to .73. This type of study assesses the degree to which different assessor teams would rate the assessee the same

TABLE 20

Assessment Center Reliability Research

STUDY	CENTER COMPONENT	TYPE OF RELIABILITY	CORRELATIONS
Glaser, et al. (1958)	situations	alternate forms	.34 - .74
Albrecht, et al. (1964)	OAR[1] (4)	inter-rater	.04 - .74
Dicken and Black (1965)	dimensions	inter-rater	.68 - .99
Bray and Grant (1966)	situations	inter-rater	.60 - .75
	dimensions	inter-rater	.31 - .54
	interview	inter-rater	.92
Greenwood and McNamara (1967)	situations	inter-rater	.60 - .75
Grant and Bray (1969)	dimensions (from interview)	inter-rater	$\bar{x}$ = .82
Thomson (1970)	dimensions	inter-rater	.74 - .93
	psych. vs mgrs.	inter-rater	.85 vs .89
Finley (1970)[2]	dimensions	inter-rater	.19 - .65
Carleton (1970)[2]	dimensions	inter-rater	.01 - .65
McConnell and Parker (1972)	dimensions	internal consistency	.64 - .90
	OAR	internal consistency	.85 - .98
	OAR	test-retest	.74
Moses (1973)	dimensions	test-retest	.49 - .72
	OAR	test-retest	.73
Bray, et al. (1974)[3]	situations	test-retest	.06 - .33
	dimensions	test-retest	.06 - .71
Hinrichs and Haanpera (1976)	dimensions	internal consistency	-.04 - .73 ($\bar{x}$ = .49)
Dyer and Hilligloss (1977)	situations	inter-rater	.70 - .80
Silzer (1978)	dimensions	inter-rater	.10 - .88 ($\bar{x}$ = .54)
Parker (1978)	dimensions	internal consistency	.84 - .94
	OAR	internal consistency	.94
Clingenpeel (1979)	dimensions	inter-rater	.53 - .72
Sackett and Hakel (1979)	dimensions	inter-rater	.42 - .88 ($\bar{x}$ = .69)

---

[1] OAR = overall rating

[2] Used supervisory ratings as comparison which are, in effect, criterion measures.

[3] This based on 8-year retest. One might reasonably expect some development to have occurred in the interim, thereby lowering these estimates.

way. In comparison with other psychometric procedures, this is fairly acceptable; nevertheless, there is clearly room for improvement as well as a need for further studies to establish reasonable standards for the assessment center field.

Third, the two studies based on the internal consistency measure of reliability show somewhat contrasting results. The research of Parker (1978) and McConnell and Parker (1972) on the AMA program displays quite high consistency for the dimensional ratings. Hinrichs and Haanpera (1976), in contrast, find much less consistency (particularly for four dimensions). This latter research suggests that certain dimensions are simply not consistently tapped in each exercise. Therefore, one must either redesign the performance situation or stop forcing the assessors to make judgments on dimensions for certain situations. Unfortunately, one might find that certain dimensions are, in fact, multi-dimensional such that different situations "pull for" different aspects of the dimension. In this case, the redesign of the dimension is necessary. Much greater research needs to be done on the issue of internal consistency since this measure provides greater clarity with respect to what the center is truly measuring across multiple assessment procedures.

In summary, research on reliability demonstrates a fair degree of reliability in measurement from the assessment center process, though in many cases there is considerable variation among dimensions (suggesting the need for redesign.) One cannot automatically assume reliability in the process; rather, the rigor of assessor training is crucial. Richards and Jaffee (1972), for example, have shown that inter-rater reliability can be substantially improved through training. Sackett and Hakel (1979) have shown that assessor teams develop stable rating patterns almost in spite of the instructions to rate dimensions independently. Thus, the design and quality of assessor training is fundamental to both the reliability and validity of the ratings.

What Do Assessment Centers Really Measure?

At first glance, the answer to this question is obvious. Assessment centers are designed to measure the behavioral dimensions found through research (or simply believed) to be important human characteristics determining successful performance on the target job. However, the reality here is somewhat more complex, as was suggested in our earlier discussion of construct validity. Let us review some of the more common approaches to this issue.

One fairly common strategy employed in many of the studies cited previously is to examine the contribution of specific tests or situations, or specific dimension ratings, to the final overall rating (via simple correlation). Such studies typically find certain dimensions to be more highly correlated with the final rating than others. A related strategy is to compare the differential contributions of tests, assessor ratings, or other measures to the prediction of a criterion measure (see, for example, Slivinski, et al., 1979; or Wollowick and McNamara, 1969). Unfortunately, neither strategy really helps the observer understand the basic constructs being measured, and both strategies usually assume that dimensional or situational ratings are independent. In fact, most studies find that dimensional ratings are highly intercorrelated. Given the complexity of the assessment center process, it is often difficult to assess whether these intercorrelations are due to rater bias, to the true interrelationships of these dimensions within the personalities of the assessees, or to the possibility that the same behavior is used by the assessors to rate different dimensions. In any case, the possibility exists that the behavioral dimensions upon which most centers are based are not the basic constructs employed by assessors in observing and rating assessees.

To tackle this problem, several researchers have used the method of factor analysis to uncover the underlying factors which seem to govern assessor ratings. Table 21 summarizes 7 such studies. It is somewhat hazardous to compare factor analyses across studies, since such factors, and the labels, depend upon the dimensions and situations employed in each study, and there is considerable variation across these studies in both areas. Nevertheless, some interesting patterns are apparent. Most of the studies identify 2 to 4 factors explaining the greatest degree of variance in the data. By far the most common factors are interpersonal skills (e.g., such dimensions as leadership, impact, or human relations skills) and administrative skills (e.g., such dimensions as organizing, planning and decision making). It has been suggested that such factors might really represent the impact of different methods in the center. For example, MacKinnon (1975) speculates that the in-basket primarily measures administrative skills; group exercises, interpersonal skills; and projective tests, motivational orientation. If, in fact, these studies simply reproduce the original methods as key sources of variation, we have not clearly established the basic constructs measured. In effect, we have measured good in-basket performance, good group discussion performance, and so forth.

The basic point here, of course, is that such factor analytic studies would suggest that assessors do not really discriminate the dimension effectively; thus, the behavioral dimensions may not be the real unit of measurement in the center. Thomson (1970) found that his assessors could discriminate dimension ratings moderately well, though supervisors making criterion ratings could not. Archambeau (1979) found that dimensional ratings within performance situations were more

TABLE 21

Factor Analytic Results in Assessment Center Research

FACTOR	STUDY					
	Bray et al. 1966, 1974	Hinrichs 1969	Thomson 1970	Archambeau 1979	Huck & Bray 1976	Sackett & Hakel, 1979
Interpersonal Skills (Leadership)	✓		✓	✓	✓	✓
Administrative Skills (Organizing and Planning)	✓	✓		✓	✓	✓
Control of Feelings (Stress Resistance)	✓	✓				
Intellectual Ability or Effective Intelligence	✓				✓	
Work-Oriented Motivation	✓		✓			
Passivity (or Activity)	✓	✓				
Dependency or Need for Structure			✓			
Sensitivity					✓	
Quality of Independent Thinking			✓			

related than ratings of the same dimension across situations. He concluded that assessors did not effectively discriminate among the dimensions. Sackett and Hakel's (1979) research is particularly enlightening since they employed more modern techniques of factor analysis, and examined the stability of assessor team ratings over time (from a data base provided by AT&T). These authors concluded that the behavioral and observational information from the center was not differentiated to a great extent. Assessor teams do develop stable patterns in rating behavior that are based on reducing the number of factors to be rated to a common set of two to three factors. In short, these studies suggest that assessment centers measure primarily a construct called interpersonal skills (or leadership) and a construct called administrative skills (or planning and organizing). Unfortunately, with few studies employing performance criteria, we do not really know how well these constructs relate to behavior on the job. Much more research is needed here.

These studies bring us full circle to the theoretical issues with which Murray, the OSS staff, and others since have long struggled; namely, what are the salient dimensions that human beings can, and do, use to perceive and organize observable behavior? It is quite possible that behavior cannot be really categorized on ten or more dimensions. For example, research in the small group field (see Leary, 1957; Osgood et al., 1957; Hare, 1976; Bales, 1970, and Bales et al., 1979) has begun to suggest that three (or at most, four) dimensions are actually utilized in the perception of behavior. Bales has identified three basic dimensions in observable behavior which are typically discriminated:

Power:	dominance, activity, impact versus dependence, passivity, withdrawal
Affectivity:	likeable, friendly, positive versus hostile, unfriendly, negative
Intellectual Control:	task-oriented, analytical, problem solving versus playful, emotional, creative

(from Bales, 1970)

It would appear that assessment center research is headed towards a similar conclusion, particularly if researchers are insistent upon minimizing the number of dimensions employed. The alternative is to design more varied performance situations to measure the individual's ability to handle various types of situations. This may be very expensive and far less parsimonious than getting to the basic constructs of the individual which are truly measurable in the assessment center. As before, we need more research to develop an adequate answer to this question; in addition, we need a theoretical framework to guide this research.

Are Assessment Centers Equitable?

Given recent activities associated with fair employment practices, it is reasonable to ask whether the assessment center process demonstrates equal validity for women or minorities. Though not readily apparent from our previous discussions, the great majority of assessment center research has been conducted on white males. Fortunately, there has been some research explicitly addressed to the issue of subgroup differences in validity.

Moses and Boehm (1975) examined the consequences of assessment for 4846 women assessed between 1963 and 1971 at AT&T. The distribution of the women across the four points of the final overall rating of management potential was nearly identical to that of a sample of men assessed during the same period. The subsequent promotion rate for these women was quite high for those rated highly in the center (over 70 percent promoted), but a subtle bias favoring women may have been operative since even low-rated women tended to be promoted at least once (over 50 percent). The proportion of high-rated women receiving second promotions was equivalent to that found by Moses (1972) for males during the same period. The predictive validity coefficient for the overall rating was slightly higher for males (.44 versus .37) but not statistically different. This study provides substantial support for the equity of the assessment process.

Huck (1974) and Huck and Bray (1976) reported the results of a study comparing the assessments of white and black females later promoted to first-level supervision (n = 91 whites and 35 blacks) or not promoted (n = 241 whites and 238 blacks). The authors discovered that on several dimensions and on the overall rating whites tended to be rated slightly, but significantly, higher. Whites also tended to be rated higher on the criterion (behaviorally-anchored) performance ratings made by supervisors. Nevertheless, the validity coefficients relating the overall rating to performance ratings and ratings of future potential were not significantly different for the two groups (.41 and .35 for performance, and .59 and .54 for potential, for whites and blacks, respectively). Thus, while many of the assessments favored whites, the pattern of validity is nearly identical for both groups.

Similar results were reported in two other studies. Parker's (1978) analysis of the AMA program found race to be slightly correlated with assessment results (.15 in favor of whites) though the difference did not reach statistical significance. Clingenpeel (1979) found that women and minorities did tend to be rated somewhat lower in the GM centers; unfortunately, subgroup validities were not reported.

At this point, the data suggests that assessment centers are reasonably equitable to different groups. Nevertheless, the research base is not terribly extensive; and there is some indication that race does bear some relationship to the ratings given to assessees (even though one study shows no differential validity despite these rating differences). Far greater research is necessary to provide a definitive answer to this question, though we would speculate that the assessment center offers greater promise than many other techniques.

Are Assessment Centers Cost-Effective?

One of the more practical uses of the research on the assessment center method is to assist a decision maker in assessing the degree to which the implementation of a center is worthwhile from a cost/benefit perspective. Unfortunately, such cost/benefit analyses are extremely difficult to conduct. While it is reasonably clear that an assessment center is more expensive than interview-or test-based selection practices (average assessment center costs are typically cited as ranging from $500 to $1,500 per participant) it is not as easy to quantify the benefits of the center. Nevertheless, a certain amount of information is available on this question.

Some authors have attempted to estimate this cost/benefit ratio by assessing the dollar cost associated with an improper selection. Such estimates typically range between $50,000 and $150,000. One author suggested that in his corporation a net benefit of 4.9 million dollars could be achieved for a cost of

340,000 dollars (see Moses & Byham, 1977:257). It is very likely that such estimates will vary considerably by organization.

The Journal of Assessment Center Technology recently completed a survey of 64 assessment center users in hopes of determining some measure of cost-effectiveness. The editors asked respondents to delineate total yearly costs for operating the center (including salaries, facilities, start-up, and continuing costs). They then asked respondents to estimate the costs associated with selection failure for 1 job, and to estimate the number of employees prevented from failure by non-selection through assessment (the product of these 2 estimates was used to estimate dollars of savings). Their findings were that average costs for a center were 24 percent of the average savings (across these organizations) and the average return on investment was 313 percent. These are impressive figures, albeit based on admittedly very rough estimates. Needless to say, it would be highly desirable to see even more rigorous research conducted on this issue.

Greater rigor is demonstrated in a recent article by Cascio and Silby (1979). These authors attempt to employ some fairly complex mathematical equations to compute the utility of an assessment center (defined as the expected gain or loss from a selection decision). Six parameters were examined: (1) the validity of the assessment center; (2) the cost of the assessment center; (3) the validity of the ordinary selection procedure; (4) the proportion of applicants actually selected for a job (i.e., the selection ratio); (5) the standard deviation of the criterion measure (in dollars); and (6) the number of assessment centers utilized. Payoff was then compared to ordinary selection and random selection as these parameters were varied.

By estimating values for each of these parameters (which admittedly can vary considerably among organizations), the authors were able to demonstrate that an ordinary selection procedure, with a validity of .25, would represent a hypothetical increase in benefits of $350,000 over random selection (assuming a selection ratio of 50 percent from a pool of 100 candidates). An assessment center at the same level of validity would produce similar savings over random selection. Nevertheless, the added value of the assessment center over ordinary selection does not show positive benefit until the validity reaches .30 (over $70,000 dollars). Each higher level of validity produces even greater benefits for the organization. Further, it is important to note that the criterion measure employed in this analysis was total yearly dollars in sales attributable to a second-line manager in a food-and-beverage firm. While this is a very interesting criterion, no research we know of in the assessment center field has actually employed it as a criterion. It will be fascinating to see whether such cost/benefit analyses will impact on the criteria used to validate centers. This would be a very positive by-product of this type of analysis.

We believe the reader would be well advised to consult this article as representative of an additional level of sophistication which is being applied to the research on assessment centers, even though the model employed by these authors rests on some difficult assumptions and judgments about many of its parameters. Through such research, we may be able to formulate a more definitive answer to the cost/benefit issue. At this time, it would be fair to say that given the proper rigor in the design and operation of a center, one might reasonably expect a fair return on one's investment in terms of increasing chances for managerial success; nevertheless, more definitive data is required to answer this question adequately.

Conclusions

It is our hope that this review of the research foundations of the assessment center concept has increased the reader's understanding of the complexities involved in studying the dynamics of the method. Clearly, much good research has been conducted to base the concept on a firm foundation. Nevertheless, it is equally clear that much more research needs to be conducted in order to advance current practice. Let us review our basic concerns with this research base, leaving our discussion of implications for education for the next, and final, section of this report.

First, assessment centers display moderately good predictive validity over a range of studies, though this validity tends to represent the relationship of the global overall rating of the center with global measures of managerial progress or ratings of future potential. Too little is known about the validity of predicting actual job performance (particularly, in lieu of the more problematic managerial ratings of performance). Given the emphasis on assessing observable behavior in the center, the use of more remote measures of performance on the job is somewhat of an anomaly. The existent data bases and ongoing research programs should help address this issue in the future. It is plainly apparent that the question of assessment center validity has not been completely answered.

Second, assessment centers display moderately good reliability in measurement, though the variation both within, and across, studies is somewhat troublesome. The sources of such inconsistencies must be more closely examined. conversely, the reasons for consistency must be examined, particularly given the fact that assessors often do not adequately discriminate among dimensions. High reliabilities may be obtained simply because assessors are rating only one or two characteristics, albeit with some stability. Closer attention to reliability issues will have a great impact on assessor training as well as the design of performance situations and observational techniques. Much more work needs to be done here to advance the current state of the art.

Third, and as a direct consequence of the second point, there is a certain lack of clarity regarding what aspects of the individual are truly measured in the assessment center process. Factor analytic studies are not conclusive evidence, but it would appear that most centers actually measure a far smaller number of dimensions of individual behavior than is apparent in the long lists of dimensions typically employed in practice. Further construct validation studies are sorely needed in order to best answer this question.

Fourth, there is limited evidence available to suggest that assessment centers do not inherently discriminate against women or minorities. Given the face and content validity of the typical center, we would expect the assessment center to be an advance over more traditional procedures. Though two studies show no differences in the selection consequences of a center, three studies do, in fact, show that white participants tend to be rated marginally higher in the center. We need more definitive analyses to determine whether this effect is due more to assessors or assessees, and, more practically, what can be done to prevent bias from operating. This is particularly important if Hinrichs' (1978) argument that assessors factor the norms of the organization into their judgments of assessees contains some merit. We suspect that this issue will receive more public scrutiny in the decade to come.

Finally, despite the limited eveidence available, it would appear that assessment centers have, at least, the potential to be highly cost-effective in comparison with random or traditional selection procedures. This is an aspect of assessment center research which is only at the earliest stages of development; however, given the proper rigor, it is an approach which could significantly advance measurement practices in the current state of the art.

V. IMPLICATIONS FOR THE EDUCATIONAL COMMUNITY

In this document we have reviewed the basic tenets of the assessment center concept, the historical context surrounding the development of the concept, the current state of the art in practice, and the foundations of the concept in empirical research. Throughout, we have attempted to describe the complexity of the issues involved in each of these areas, and we have offered our own interpretations and conclusions regarding these issues. In this final section, we wish to address the educational community more directly with respect to the key issues underlying the decision to implement the assessment center concept in educational settings.

The Assessment Center Concept: Measuring Occupational Competence?

We have shown that the assessment center concept was created as the result of the practical needs of large bureaucracies in military and business settings for more effective selection of personnel, and as the result of more theoretically inclined research attempts to measure systematically the basic components of the individual personality. The concept has evolved, through a considerable amount of practical application and research, into a complex and sophisticated assessment technology. The framework of this technology is based on the fundamental assumption that the best prediction of occupational success derives from multiple measures of multiple human characteristics within the context of simulated situations from work settings. As a concept, the assessment center undoubtedly represents a significant advancement in assessment and selection technology, and has clearly contributed to the advancement of the state of the practice in employment decision making. By virtue of this fact alone, the concept is worthy of careful consideration by the educational community.

Nevertheless, one must ask whether the assessment center concept, as currently applied in noneducational settings, is fundamentally directed towards the measurement of occupational competence. We would conclude from our review of the state of the art that this is only partially the case. Let us elaborate this conclusion in more detail.

First, such a judgment must ultimately rest on the definition of competence that one employs. The assessment center concept, as a tradition of research and practice, has developed quite apart from the tradition of competency-based assessment. This latter tradition advocates a definition of competence as a set of generic human characteristics underlying effective performance. From this perspective, an individual competency is an underlying construct of the individual that can be manifest, in alternative forms, in observable behavior. In contrast, assessment center practitioners typically subscribe to this definition in a limited fashion. Since the early attacks on Murray's formulations in the OSS program, assessment center practitioners have avoided the articulation of underying constructs of the individual. Emphasis is usually placed on observable behavior patterns as the primary constructs. Nevertheless, these behaviors must be categorized and rated; hence, the articulation of behavioral dimensions for measurement. As we have noted, these dimensions are often a mixture of observable behaviors and inferences regarding underlying skills or abilities. In short, it is not clear at this point in time whether one can, or should, equate the concept of a behavioral dimension with the concept of a competency. The two concepts are undoubtedly related, though few practitioners would feel comfortable describing these behavioral dimensions as underlying constructs (particularly given the pressure exerted by the EEOC guidelines).

Second, the behavioral dimensions are usually derived from some form of job analysis (in the ideal case). Quite often the dimensions are conceptualized, as a result, to represent the basic demands of the job (e.g., planning, decision making, leading). In observing behavior in an assessment center, therefore, an assessor must inevitably make the subtle inference that the behaviors exhibited by an assessee represent some ability to handle these job demands (e.g., an ability to plan, make decisions, or lead). This is a subtle, yet extremely critical, transformation of the data (or dimensions) from a job analysis toward a competency-based framework. Furthermore, it opens the possibility that the construct which guides the assessor's ratings is not equivalent to the definition of the dimension. For example, two assessors may rate the same assessee differently on the dimension of decision making because one uses a (hidden) construct of "degree of logic" employed by the assessee and the other uses a construct of "degree of problem conceptualization" by the assessee. An even worse problem is that the assessor may use a construct totally unrelated to the dimension at hand to reach a rating decision (this is Hinrich's [1978] basic argument). The important point here is that the behavioral dimensions employed by the typical assessment center may, or may not, represent dimensions of human competence when rated by assessors.

Third, as we examined the research base, we noted that a key problem area was the construct validity of the assessment center concept. The measured validity of the concept has been predicated primarily on the relationship of the overall rating from the center with a global measure of managerial success--typically defined in terms of upward mobility. Much of the research fails to validate the behavioral dimensions within the context of on-the-job performance. As a consequence, it is not clear that the dimensions rated in the center are related to behaviors exhibited on the job, even though this relationship is often assumed if the individual is promoted. Furthermore, the available research on the construct validity of the behavioral dimensions reaches problematic conclusions with respect to the degree to which assessors can truly distinguish the dimensions. It may well be that the typical assessment center really measures only two to four underlying constructs of the individual--and these may be highly related to the methods of measurement employed. In short, while assessment centers may be used to predict managerial success in a global fashion, they are much less precise in pinpointing the underlying human characteristics, or competencies, which underlie the job behaviors that determine success.

In summary, we reach the conclusion that the assessment center concept, as currently practiced, is at best an indirect measure of occupational competence; is most likely a measure of global potential for upward mobility; and is at worst an elaborate "policy-capturing" device by which managers reach a selection decision they would have reached anyway without the behavioral dimension ratings. Nevertheless, there is little within the present conceptual framework or technology which would inhibit the more direct measurement of occupational competence. One only needs a basic framework for describing competencies as the result of a job analysis, a research design oriented to measuring these competencies explicitly within the center and on the job, and a testable theory to relate these competencies to performance and to success. This is not an easy task, to be sure, but it would represent the combination of two traditions that could significantly advance the current state of the art in assessment of occupational competence.

Obviously, we believe that determining what an assessment center truly measures is an important issue per se. However, we raise this issue as a fundamental concern with the current state of the art because of its implications for educational applications. The assessment center concept has been developed over the years primarily within business settings (and at a later stage, governmental settings) as a tool for selection. As a consequence, practitioners and researchers have been more concerned with the validity of the overall rating (i.e., the selection decision in most cases) than with the validity of the behavioral dimensions underlying this overall rating.

With the late entry into the educational arena of assessment center utilization (in the last five to seven years), and with the increase in the number of developmental applications in noneducational settings, the emphasis in measurement is slowly beginning to shift towards more precision in the definition and assessment of the specific behavioral dimensions. The need for this shift in emphasis is dictated by the demand to provide assessees with reliable and valid assessments of their strengths and weaknesses. It is likely that this has been, and will continue to be, of fundamental importance to educational applications--particularly at the level of student assessment (the selection of faculty and administrative personnel may be more closely related to past practice in other settings). In short, it is reasonable to conjecture that educators will be less interested in predicting student success in the world of work (though such predictions are by no means unimportant in the long run) and more interested that the assessment center provides reliable and valid measures of specific competencies as input to the student's learning process and progress.

From this perspective, the current state of the art might be described as fairly advanced for selection decisions, but at a very early stage for applications to development. Thus, the educator must be aware that extension of the assessment center concept to educational applications will necessarily entail further refinement of the technology. Much of the previous research base may simply not be directly applicable. The need will exist to conduct research on these applications in order to assess more directly the adequacy of these competency measurements. We suspect that many educatonal institutions will not care to take on this burden; for those who do, it is likely that the solutions generated will contribute substantially to the state of the art. The educational arena may well be a prime setting for the future development of the assessment center concept within a competency-based framework.

Assessment Centers in Education: Is There a Need?

The preceding discussion has focused primarily on extension of the concept of the assessment center, as a potential method for measuring occupational competence, to educational settings. Of equal importance to the educator is assessing the degree to which the institution really needs this technology in relationship to its own goals and purposes. Certainly the number of educational applications is increasing and the number of advocates is proliferating. But the educator must cautiously assess his or her temptation to "jump on the bandwagon." Is there a reasonable need for the application which will justify the energy and resources necessary to implement the center? Let us enumerate some possible needs facing the typical educational institution for which an assessment center might be considered a solution.

First, the simple fact that assessment centers are being increasingly utilized by employers to make selection or development decisions should be cause for some consideration by educators. It is likely that more graduates of the educational system will one day experience an assessment center evaluation as the vehicle for attaining further employment or training opportunities. It is conceivable that educators will face some demands from students (or even faculty) to provide educational experiences with assessment centers, or at a minimum, to provide counsel and assistance on how to prepare for such an assessment. (This could be a compelling pressure if professional certification agencies turn to the use of assessment centers for licensing and certification decisions.) We would guess that such pressures would be of low intensity and would require fewer resources for an adequate response. Nevertheless, it is possible that the educator may discover the need to examine more closely the content of the curriculum and the process of learning in the context of the assessment center concept.

This line of reasoning leads to a second, and more substantive, consideration. We have shown that the assessment center concept is based on the development, through a more or less rigorous job analysis, of a set of behavioral dimensions believed to represent the individual behaviors predictive of success in the employing organization. While it is debatable whether these dimensions represent underlying competencies of the individual, it is clear that employers using the assessment center technology are attempting to be far more specific, and rigorous, in pinpointing the individual determinants of success --as success is currently defined. We believe that educational institutions have an important interest in these definitions of success and in the specification of the human characteristics presumed to be underlying success.

The traditional view of the function of the educational institution is as a retainer and purveyor of general and specific knowledge at the base of our culture and society. We would not devalue this perspective; however, the educational system can also be seen as functioning to provide competent human resources for the world of work. From the viewpoint of the student, the educational process should better enable him or her to identify, cope with, and achieve success in the opportunity structure of the larger society. If employers continue to use more sophisticated definitions and measurements of the basic human characteristics they believe to predict success in their opportunity structure, then educators may face pressures to define more clearly the outcomes of the educational process in similar terms. Stated simply, by ignoring the development of assessment center technology (as one method for assessing occupational competence), the educational community may lose the initiative, to other sectors of our society, for defining the true value and results of the educational experience.

This raises a third, and more general, consideration. It would appear that educators are facing a new age of accountability. One need only cite such problems as the spiraling costs of education in relationship to the benefits produced for the individual; the demand from minority groups or other parties to the educational system that the educational process increase access to the opportunities of our society; the demand by various public groups that education produce measurable outcomes in student achievement; or the fact that simple possession of a diploma, in an era of "grade inflation," is no longer compelling evidence of competence for many employers. In short, educators may be faced with several sources of external pressure to document more precisely the benefits and outcomes of the educational experience. If the educational community seriously begins to address these issues, we would speculate that the assessment center concept will be worthy of consideration. Nevertheless, the educator must be prepared to question the entire educational process--from curriculum design to teaching process and assessment procedures--for the assessment center concept places emphasis on what the individual can do and not on what he or she knows. In this sense, the concept represents a real challenge to the educational community to document rigorously the outcomes of the educational process for the individual.

Fourth, educational institutions, like other organizations, are continually required to assess the performance of their members--be they administrators, faculty, or students. Thus, despite the external pressures emanating from the society at large, educators face internal needs for accurate assessment. This review would suggest that the assessment center concept can, at least in theory, be applied to these activities, though the various applications will likely encounter different problems.

For example, the selection and/or development of administrative personnel--particularly from the faculty ranks--is a common problem cited by educators. We would estimate that the transfer of assessment center technology from business settings to this type of application in educational settings would be less problematic with respect to implementation than other types of applications, simply because this is the most common application in current practice in other settings. Certainly the educator would need to attend to the host of technical problems cited in the conclusion of Section III of this report with respect to the design and operation of the center. Nevertheless, with the proper expertise applied, it is probable that most educational institutions could substantially improve practice in this area through implementation of an assessment center process. The educator must fundamentally consider the true extent of the need. This consideration will likely involve the number of vacancies within the administrative ranks in any given time period; past patterns in advancement rates; the degree to which administrative incompetence is problematic; the degree to which the institution can accommodate to a more public set of criteria for the selection process; and the degree to which the selection criteria (i.e., behavioral dimensions) can become part of the ongoing appraisal system for these personnel.

The selection and/or development of faculty via an assessment center process represents, we would judge, a slightly higher risk. This type of application is at present very rare. The basic problem here, of course, is the tradition apparent in many educational institutions (especially at the postsecondary level) for the selection of new faculty by faculty committees within disciplines. The selection criteria typically employed are not very specific, and tend to rest upon such factors as academic credentials and publications. The implementation of an assessment center process for the purpose of faculty selection would probably engender great debate on the criteria to be utilized for evaluation, both within and across disciplines. Nevertheless, this exercise may be quite valuable from the viewpoint of the educational administrator, and could serve to clarify the institution's objectives for faculty competence. If the educator is prepared to encounter the resistance which we would anticipate, it is likely that an assessment center could contribute to the improvement of the teaching process via faculty improvement. (For example, a conceivable application is the design of a developmental center in which incumbent faculty are assessed in situations such as student counseling and lecture delivery, with the aim to improve skills in these areas. The key here, once again, is conducting the research necessary to identify the competencies underlying effective performance in these situations. Therein might lie the great debate.)

The application of the assessment center concept to the student body of an educational institution is presently far more common and can potentially have the farthest reaching implications for both the content and process of education.

Many present educational applications have been designed to provide developmental feedback to students and to provide more guidance in student counseling with respect to career choices or skill development within specific careers. It is likely, in our judgment, that such applications will continue to be the most common. They are easier to design; many suppliers offer packaged programs to educational institutions for such use in counseling; and this type of application is less problematic with respect to the design of other aspects of the institution (e.g., curriculum, course evaluation, or criteria for graduation). Certainly, the assessment center utilized as a device for student counseling can make a substantial contribution to the quality and effectiveness of that counseling. Nevertheless, as we have concluded at several points in this document, the current state of the art needs to be considerably advanced so that such counseling is based on valid and reliable assessments of internally consistent constructs of individual competence. It is our hope that these applications include the research necessary to validate these constructs. Otherwise, practice will continue to outstrip the theoretical and empirical base.

Other applications to the student body are conceivable, though much more troublesome for implementation. It is a logical step to extend the assessment center concept to initial student recruitment and selection at the postsecondary level. If undertaken, educators would be forced to examine more closely the role of the student, and to establish more clearly criteria of student success. If these criteria were merely grade-point averages at the conclusion of college, it is unlikely the costs involved would weigh favorably over simple, entrance (paper-and-pencil) examinations. However, if an educational institution were seriously interested in predicting success as defined by an increase in basic competencies (such as particular cognitive or interpersonal skills) of the student, an assessment center might prove to be more useful.

By far the most radical application would be the use of the technology to evaluate student performance with respect to advancement within the institution or with respect to granting of the certificate of graduation. We have already alluded to the external pressures which might lead an educational institution in this direction. If an educator is at all concerned with documenting the impact of the educational process on the growth of the student (in terms of actionable capacities or abilities beyond the simple possession of knowledge), it is likely that the assessment center concept can make a substantial contribution. Nevertheless, such an application will require extensive collaborative efforts by faculty, administration, and students to weather the truly monumental change in the institution represented by this implementation--witness the efforts at Alverno College where such an applicaton is in process.

Educators would need to create the vehicle for a collective definition of the outcomes of the educational process; and these outcomes would need to be defined in terms which are specific with respect to observable indicators yet generic to the extent that they are not discipline-specific. The design of the assessment techniques would require careful examination as would the training of faculty assessors. Students would require some re-orientation to the new criteria for evaluation. A commitment to the research validity of these constructs of assessment would be necessary to ensure that students were, indeed, truly being measured on dimensions of competence specified by the institution. Finally, the entire institution would need to consider the implications of the assessment procedures for the design of the curriculum content as well as for the design of the teaching process. All of this would have to take place within the context of a willingness to innovate, experiment, and change, since this application clearly exceeds current practice in other settings. Without this desire, or an underlying institutional consensus to move in the direction of outcome-based measurement, the educator may find this application to be exceedingly high-risk.

In summary, it is clear that educational institutions can employ the assessment center concept to great benefit at a variety of levels of involvement. Depending upon the nature of the application, the educator can also expect differing types of problems. The most important point to consider is the degree to which the application strikes at the heart of evaluating the educational process. While there is always a need for improving these evaluations, the use of assessment center technology may represent a "quantum jump" for the institution. This is a risky proposition at best, yet the payoffs may be equally high for improving the educational process. We would simply caution the educator to be aware of the difficulties associated with implementation before proceeding down the road.

The State of Practice: A Problem for Education?

We have shown that the number of applications of the assessment center concept has rapidly expanded in the past ten years. Several observers of this phenomenon fear that the present faddishness of the concept may be quite detrimental to future development, and may seriously weaken the research base of current practice. We strongly concur with this opinion. We believe that this problem is particularly critical for the educational community since applications in this arena are of much more recent origin. Far less is known about the true strengths and weaknesses of these applications. Despite the growing need and desire to implement assessment centers in educational settings, educators must approach assessment center implementation with extreme caution and with lots of help. Let us pinpoint the aspects of the technology which may be particularly problematic for educators.

First, the wisdom of current practice dictates that the behavioral dimensions rated in an assessment center should be derived from some form of job analysis. Unfortunately, there is some debate within current practice over the best method for conducting such analyses. The most common practice is to categorize job demands and job functions, the result being that the behavioral dimensions represent behaviors required by the job. As we have argued previously, this approach is not equivalent to documenting the individual competencies which enable the individual to perform the job in a superior fashion. Nevertheless, when one examines the dimensions often employed in current practice, there is still some confusion evident between the former _job-related_ approach and the latter _person-related_ (or competency-based) approach.

What is clear is that the design of educational applications for students will most often require _person-related_ definitions of the behavioral dimensions. This is true to the extent that the institution produces graduates for a wide array of jobs and not for one type of job or profession. Thus, the institution will need to identify more generic skills (or competencies) that enable an individual to be successful in a variety of work settings. The alternative response, and one which we noted in current practice, is to simply expand the list of dimensions to the extreme (over 60 in one case) in order to "cover the bases." Unfortunately, present research in other settings suggests that assessors simply cannot evaluate an individual on so many dimensions.

In short, current practice offers few guidelines for the educational community on how best to derive the behavioral dimensions to be rated in an assessment center oriented towards student evaluation. What is missing is a theoretical framework for the description of human competencies in relation to the world of work. To a certain extent, the educational community is faced with the same problems faced, but not totally resolved, by the early work of the OSS and those studies appearing after World War II (including Bray's Management Progress Study). Not much work has occurred on this front in the intervening years. Nevertheless, if current practice is to expand in the educational arena, the issue of behavioral dimension definition (and conceptualization) must be confronted directly. This will likely be both an exciting opportunity and also a frustrating problem for the educational community.

Second, the design of performance situations and other assessment techniques employed in the typical assessment center can certainly be the most creative aspect of implementation. We noted in this review that the typical center employs four fairly common techniques (the in-basket, the leaderless group discussion, the role play, and the interview). Current practice has led to the development of a variety of other assessment formats. The most important issue at this point is assessing what aspects of individual competence are best assessed by what types of techniques. We believe that too much energy has been devoted to demonstrating the validity of assessment centers relative to that of other techniques, and too little energy has been devoted to the determination of what different techniques actually measure, specifically or in common with each other. Fortunately, more research is being devoted to this issue at present. Nevertheless, educators will face some difficult decisions in determining which techniques to include in an application. Our judgment is that educational applications should include a wider array of techniques (such as was embodied in the original OSS program) and should include the construct validation research necesary to disentangle the best measures for specific competencies. Without such efforts, we may never really know how cost-effective different techniques are in relationship to their measurement quality.

Third, it is eminently clear that the effectiveness of the assessment center concept ultimately hinges on the quality of the assessor observations and ratings. It is equally evident that this is an area of the current state of the art that is not adequately researched or developed. While the problems here are certainly not insurmountable, it is likely that educators will face many difficult challenges in the selection of assessors, in the design of assessor training, and in the design of the dimension rating scales. Such issues, once again, will be particularly critical when the assessment center is applied to student evaluation, since the assessors may well be faculty from different disciplines who do not share a common framework for evaluating student performance. We noted in this review that many present educational applications do not appear to place much emphasis on assessor training or on the design of observation and rating formats. Our judgment is that this represents an extreme weakness in current practice with potentially disastrous consequences for the reliability and validity of the assessments. Educators will be wasting their money if such practices are allowed to continue.

Fourth, we have shown that the average assessment center typically produces overall ratings on each of the dimensions as well as a final overall rating which represents some combination of the dimensional ratings and/or a prediction of future success. While we have labelled these ratings as "competency judgments," it is evident from our prior discussion that this label may be something of a misnomer. Certainly there is considerable debate in current practice over how these final ratings ought to be obtained (i.e., by statistical aggregation or by clinical judgment and discussion of the assessor team), and there is a certain lack of clarity as to what these overall ratings truly represent in concept. On the one hand, this problem may be of little concern to educators if the purpose of the application is to profile strengths and weaknesses of the individual. Here, of course, much more concern needs to be devoted to the precise definition and measurement of each of the dimensions. On the other hand, if the application does involve multiple measures of each dimension, the educator will be faced with deciding how to summarize the assessments adequately. If the application is for student evaluation (for advancement or for graduation), this issue may become extremely important. Unfortunately, at present there are no clear guidelines available for resolving this problem, and much effort will have to be devoted to this problem within the application per se. The resolution will clearly indicate the parameters necessary for adequate feedback to assessees.

Finally, the current state of practice is replete with unsubstantiated positive claims of benefit and negative charges of harm for both the individual assessees and the organization as a whole. The important point for the educational community is that each of these speculations must be examined closely both prior to implementation of a center and as part of the implementation effort. We noted several such consequences in Section III of this report. The educator would do well to consider these consequences closely, and devise means to measure their impact in any implementation effort.

In summary, we would conclude that the current state of practice does present some difficult, though not insurmountable, obstacles to implementation in the educational context. Any implementation effort will require close examination of these obstacles at the earliest stages of design and a coherent plan of action to overcome them. Educators will need to gather the best resource expertise available, both from within and outside of the institution, in order to solve these problems. Such efforts could lead to a considerable advance in current practice; in their absence, we fear that future practice will create more questions than answers, and will ultimately dilute the effectiveness of the assessment center concept in the educational context.

The Research Base: Further Research Needed?

It is somewhat painful to conclude this document with the researcher's ultimate caveat that further research is needed; nevertheless, it is obvious from our review of the state of the art that while much good research has been conducted, far too many questions are at present unanswered in the research base. The educational community must treat claims of the "proven validity" of the assessment center concept with a degree of skepticism. We have attempted to identify the complex issues involved in this arena of the current state of the art. Let us reiterate our research priorities in relationship to the educational context.

First, the most important research issue from the viewpoint of education is, in our judgment, the determination of what an assessment center truly measures through focused construct validation research. The many criterion-related studies which have been reported are certainly important, but the choice of global criteria of advancement, and global predictors from the assessment center process, has provided little information on the actual competencies being measured, or on how these competencies are reflected in actual job performance. Since we believe that educators will be more concerned with measuring generic competencies for the world of work, we are convinced that more research will be required on this problem in order to advance educational applications. Such research, of course, would have tremendous implications for the design of behavioral dimensions, the selection of alternative assessment techniques within the center, and the design of observational formats. In effect, the assessment center concept should probably no longer be treated as a unitary technique in comparison with other techniques; rather, it should be treated as a true multiple-measure technology with varying components dependent upon the nature of the human competencies one wishes to measure.

Second, within educational contexts, we expect a great concern will be expressed for how the educational process can influence growth in competence as evaluated via the assessment center process. Virtually no research has been done on this question. Nevertheless, the growing trend for use of assessment for developmental purposes will demand that researchers examine more closely how the behavioral dimensions evaluated in the center can be influenced by education and training.

Third, if research can identify the specific constructs measured in the center, and can relate these constructs to the educational influences on the individual, then we can ask the more criterion-related question of how competencies appear in job performance and contribute to the individual's success in the world of work. For most educational institutions, this research would represent an ambitious undertaking requiring longitudinal research designs spanning many years. Nonetheless, such designs would clearly contribute to the debate on assessing the real value of the educational experience.

Fourth, we have noted that the current state of the art is just now reaching the level of sophistication necessary to rigorously document the cost-effectiveness of assessment center implementation. In an era of tightening resources in the educational arena, such data are badly needed by decision makers. It is our hunch that such cost-benefit analyses may need to be more refined than they are at present; that is, educators may require a model for cost-benefit analysis which assesses the utility of alternative assessment techniques within the center for measuring specific competencies. Such research would undoubtedly be difficult but probably more informative to the educational community than that which is already available in comparing the costs and benefits of paper-and-pencil tests, simulations, and the like.

Fifth, there are a host of more technically-oriented research issues which need attention in the practical design of a center. Much of this type of research could be conducted experimentally or as part of an application. Such issues could include, for example:

- Determining the impact of various numbers, and types, of behavioral dimensions on assessor ratings;
- Determining the critical aspects of design for role plays and simulations;
- Determining the impact of various assessor team compositions on ratings generated;
- Determining the impact of alternative observational formats and associated assessor training programs;
- Determining the impact of assessor team dynamics on the quality of the ratings generated; and
- Determining the impact of alternative forms of feedback on assessees.

Such a list could be endless, but obtaining answers to these technical problems would serve to improve present standards in many areas of current practice.

Finally, given the pressure from governmental agencies and the courts to ensure equity in testing and employment decision making, the assessment center has emerged as a viable alternative to other assessment practices. The present research base is promising, though not conclusive, regarding the equity of assessment. Research within educational institutions on this issue could be done easily and could contribute valuable data to this public dialogue.

In summary, we would conclude that the research foundations of the assessment center concept provide some justification for extension of applications to educational settings; however, the need for further research of a quite different focus is equally evident. Without such research as has been identified in this report, the assessment center concept will not fulfill its promise as an assessment technology. Educators should not proceed with an application without including an associated research design to answer at least some of the heretofore unanswered questions evident in the current state of the art.

APPENDIX:

Standardized Performance Situations

Standardized Performance Situations

In-Basket

An assessee is told he or she has just been promoted to a management job. His or her predecessor vacated the job unexpectedly without cleaning up a number of matters filed in the in-basket. The assessee must go into the office on a Sunday and deal with the accumulated memos, letters, and reports. No one is in the office to help the assessee who must decide which requests and tasks are most important, which items can be deferred for further information, and which must be acted upon. The assessee typically has 1-1/2 or 2 hours to deal with the entire contents of the in-basket, and to write down every action he or she would take (including notes, letters, meeting agenda, and phone calls.)

In-Basket Interview

Following the in-basket exercise, each assessee is interviewed by an assessor who sorts through the product, making sure each action is clear. The assessor asks for the rationale behind actions taken, the approach the assessee took to organize the task, and any other information that might furnish insight into the assessee's skills vis-à-vis decision making, organizing, and other administrative tasks.

Structured Interviews

The interview typically lasts from 1 to 2 hours and is conducted by an asssessor. The degree of structure varies from interviews of prescribed questions (such as an oral questionnaire) to a suggested list of discussion topics. Questions are used to elicit information on specific knowledge, skills, or abilities evaluated in the assessment center. The interview may be used to access information not available by other methods or may be used to confirm inferences drawn from behavior in other assessment exercises. Often included as one of the first experiences the assessee undergoes in the process of assessment, the interview can be used to answer questions and assuage the participant's apprehensions.

Business Games

Business games are individual or team exercises which simulate a business process and require the assessees to diagnose, plan, and act in the face of given market and resource parameters. For example, the assessee may have to purchase toy parts to fabricate a product and then have to sell it. Prices of materials and market conditions may change. The assessee must make choices about building or disposing of inventory product mix, rate of production, pricing, and the like.

Fact Finding

In a fact-finding exercise, an individual (or a team) may be presented with a problem to be resolved in a specified period of time. The assessee is not given adequate information to resolve the problem and, therefore, must seek out information useful for making a decision. Information may be available from informed sources (i.e., assessors prompted to give only certain information if it is specifically requested), written material, or other assessees (they may all be given different pieces of information needed to solve the problem). After the allotted time, the assessee must make his or her recommendation for a solution and is evaluated on the thoroughness of the search and the quality of the presented solution.

Leaderless Group Discussions

The most common format for this exercise calls for the assessees to be assigned different roles in a decision-making group or advisory council. After each is given information which details a different position, they are then required to advocate their positions in a leaderless group meeting from which only one position will emerge as the recommendation of the group. This advocacy paradigm requires the assessees to be competitive. The competitive atmosphere elicits a variety of opportunities for demonstrating persuasiveness, resistance to stress, forcefulness, willingness to compromise, and other characteristics.

A second type of leaderless group discussion is based on a situation in which group members do not have assigned roles and must cooperate to solve a mutual problem. Behaviors such as openness to others' ideas, emergent leadership, and organization can be observed in these types of groups.

While competitive or cooperative groups may be formed with or without assigned roles, the preceding two combinations are the most common. Each combination of role/nonrole assignment and cooperative/competitive purpose may elicit different kinds of strategies and provide opportunities to observe different dimensions of job performance.

REFERENCES

Alamprese, J. *New York State External High School Diploma Program: Joint Dissemination Review Panel Submission*, 1979.

Albrecht, P. A., Glaser, E. M., & Marks, J. Validation of a multiple-assessment procedure for managerial personnel. *Journal of Applied Psychology*, 1964, *48*(6), 351-360.

The AMA Assessment Center Program: Assessor's Guide. New York: AMACOM, 1975.

Archambeau, D. J. Relationships among skill ratings assigned in an assessment center. *Journal of Assessment Center Technology*, 1979, *2*,(1),7.

Assessment and Development. A monthly newletter published at Pittsburgh: Development Dimensions International.

Baehr, M. E. Job analysis: Objectives, approaches, and applications. In E. A. Miller, E. H. Burack, & M. Albrecht (Eds.), *Human resource management readings*. Englewood Cliffs: Prentice Hall, 1979.

Bales, R. F. *Personality and interpersonal behavior*. New York: Holt, Rinehart and Winston, 1970.

Bales, R. F., Cohen, S. P., & Williamson, S. A. *Symlog: A system for multiple level observation of groups*. New York: The Free Press, 1979.

Barrett, R. S. *Statistical diagrams*. Publisher unknown, 1978.

Barrett, R. S. *Zetetic for testers II*. Publisher unknown, 1978.

Bass, B. The leaderless group discussion. *Psychological Bulletin*, 1954, *51*, 465, 492.

Bender, J. M. What is "typical" of assessment centers? *Personnel*, July-August 1973, 51-57.

Bentz, V. J. Validity studies at Sears. In W. C. Byham (Chair), *Validity of assessment centers*. Symposium presented at the meeting of the American Psychological Association, Washington, DC, 1971.

Bentz, V. J. The Sears' experience in the investigation, description, and prediction of executive behavior. In F. R. Wickert & D. E. McFarland (Eds.), Measuring executive effectiveness. New York: Appleton-Century-Crofts, 1967.

Boche, A. Management concerns about assessment centers. In J. L. Moses & W. C. Byham (Eds.), Applying the assessment center method. New York: Pergamon Press, 1977.

Boehm, V. R. Negro-White differences in validity of employment and training selection procedures: Summary of research evidence. Journal of Applied Psychology, 1972, 56, 33-39.

Boehm, V. R., & Hoyle, D. F. Assessment and management development. In J. L. Moses & W. C. Byham (Eds.), Applying the assessment center method. New York: Pergamon Press, 1977.

Bray, D. W. The assessment center method of appraising management potential. In J.W. Blood (Ed.), The personnel job in a changing world. New York: AMA, 1964, 225-234.

Bray, D. W. The management progress study. American Psychologist, 1964, 19, 419-420.

Bray, D. W. The assessment center method. In Craig (Ed.), Training and development handbook, 1976.

Bray, D. W. Assessment centers. Seminar offered at the Massachusetts Psychological Center, May 13, 1978.

Bray, D. W., & Campbell, R. J. Selection of salesmen by means of an assessment center. Journal of Applied Psychology, 1968, 52, 36-41.

Bray, D. W., Campbell, R. J., & Grant, D. L. Formative years in business: A long-term AT&T study of managerial lives. New York: Wiley & Sons, 1974.

Bray, D. W. & Grant, D. L. The assessment center in the measurement of potential for business management. Psychological Monographs, 1966, 80 (17, Whole No. 625).

Bray, D. W., Grant, D. L., & Campbell, R. J. Studying careers and assessing ability. In A. J. Marrow (Ed.), The failure of success. New York: AMACOM, 1972, 154-169.

Bray, D. W., & Howard, A. Keeping in touch with success. The Wharton Magazine, Winter 1979, pp. 28; 33.

Bullard, J. F. An evaluation of the assessment center approach to selecting supervisors. Corporate Personnel Department, Caterpillar Tractor Co., May 1969. (Mimeo)

Byham, W. C. Assessment centers for spotting future managers. *Harvard Business Review*, 1970, *48*(4), 150-160.

Byham, W. C. How a scientific assessment center works. In A. J. Morrow (Ed.), *The failure of success*. New York: AMACOM, 1972, 136-153.

Byham, W. C. How to improve the validity of an assessment center. *Training and Development Journal*, November 1978, 4-6.

Byham, W. C., & Huck, J. R. *Bibliography of publications: Identification of supervisory and managerial potential*. Pittsburgh, PA: DDI, 1977.

Byham, W. C., & Wettengell, C. Assessment centers for supervisors and managers: An introduction and overview. *Public Personnel Management*, 1974, *3*(5), 353-364.

Campbell, R. J., & Bray, D. W. Assessment centers: An aid in management selection. *Personnel Administration*, 1967, *30*(2), 6-13.

Campbell, J. P., Dunnette, M. D., & Hellervik, L. Development and evaluation of behaviorally based rating scales. *Journal of Applied Psychology*, 1973, *57*, 15-22.

Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, *56*, 81-105.

Career Analysis Procedure Manual. Hartford: Life Insurance Marketing and Research Assocation, 1976.

The Career Development Center: A Brief Overview. Menlo Park, CA: Zenger/Miller & Associates, Inc.

The Career Development Center Program Evaluation: November 1974-May 1976. Menlo Park, CA: Zenger/Miller & Associates, Inc., 1977.

Carleton, F. O. *Relationships between follow-up evaluations and information developed in a management assessment center*. Paper presented at the 78th annual convention of the American Psychological Association, Miami Beach, FL, 1970.

Cascio, W. F., & Selby, V. Utility of the assessment center as a selection device. *Journal of Applied Psychology*, 1979, *64*(2), 107-118.

Clingenpeel, R. E. *Validity and dynamics of a foreman selection process*. Paper presented at the International Congress on the Assessment Center Method, New Orleans, LA, June 1979.

Cohen, B. M., Moses, J. L., & Byham, W. C. *The validity of assessment centers: A literature review*. Monograph II. Pittsburgh, PA: DDI, 1977.

Cronbach, L. J. *American Psychologist*, 1955, *10*, 419.

Cronbach, L. J. *Essentials of psychological testing*. New York: Harper and Row, 1970, 667-694.

Dapra, R. A., & Byham, W. C. Applying the assessment center method to selection interviewing. *Training and Development Journal*, April 1978, *32*(4), 44-49.

DeNelsky, G. Y., & McKee, M. G. Prediction of job performance from assessment reports: Use of a modified Q-sort technique to expand predictor and criterion variance. *Journal of Applied Psychology*, 1969, *53*, 439-445.

Department of the Air Force. *Background paper on the SOS leadership skills program*. Unpublished paper.

Development Dimensions, 1977-78 Catalogue. Pittsburgh: DDI, 1977.

Development of assessment techniques to select managerial personnel. Sacramento, CA: Selection Consulting Center, January 1979.

Dicken, C. F., & Black, J. D. Predictive validity of psychometric evaluation of supervisors. *Journal of Applied Psychology*, 1965, *49*(1), 34-47.

Dicostanyo, F., & Audrietta, T. The supervisory assessment center in the Internal Revenue Service. *Training and Development Journal*, 1970, pp. 12-15.

Dodd, W. E. Attitudes toward assessment center programs. In J. L. Moses, & W. C. Byham (Eds.), *Applying the assessment center method*. New York: Pergamon Press, 1977.

Dodd, W. E. Validity studies at IBM. In W. C. Byham (Chair), *Validity of assessment centers*. APA symposium, Washington, DC, 1971.

Dodd, W. E. *Will management assessment centers insure selection of the same old types?* Paper presented at the annual convention of the American Psychological Association, Washington, DC, 1971.

Donaldson, R. J. *Validation of the internal characteristics of an industrial assessment center program using multitrait-multimethod matrix approach*. Doctoral dissertation, Case Western Reserve University, 1969.

Dyer, F. N., & Hilligloss, R. E. *Using an assessment center to predict field leadership performance of Army officers and NCOs*. Paper presented to the 19th Annual Conference of the Military Testing Association, San Antonio, TX, October 19, 1977.

Earles, J. A., & Winn, W. R. *Assessment centers: An annotated bibliography*. Personnel Research Division, Lackland AFB, May 1977.

Educational Leadership Appraisal (ELA): A Progress Report. Fort Lauderdale, FL: Nova University, 1977-78, *7*(2).

Finkle, R. D. Managerial assessment centers. In M. Dunnette (Ed.), *Handbook of industrial and organizational psychology*. Chicago: Rand-McNally, 1976, 861-885.

Finley, R. J., Jr. *An evaluation of behavior predictions from projective tests given in a management assessment center*. Paper presented at the meeting of the American Psychological Association, Miami Beach, Florida, 1970.

Fleming, J. L., & Sells, S. B. *The corporate personnel assessment center: A review and evaluation of potential and performance*. Eastern Airlines Research Training Project No. 1-225-156, August 1972.

Frank, F. D., & Whipple, D. An assessor certification program based on simulation of the assessor job. *Journal of Assessment Center Technology*, 1978, *1*(2), 1-5.

Frederiksen, N., Saunders, D. R., & Wand, B. The In-Basket Test. *Psychological Monographs: General and Applied*, 1957, *71*(9), 1-28.

Garforth, F. I. War office selection boards. *Occupational Psychology*, 1945, *19*, 97-108.

Ghiselli, E. E. Dimensional problems of criteria. *Journal of Applied Psychology*, 1956, *40*, 1-4.

Ghiselli, E. E. The validity of aptitude tests in personnel selection. *Personnel Psychology*, 1973, *26*(4), 461-477.

Ginsberg, L. R., & Silverman, A. The leaders of tomorrow: Their identification and development. *Personnel Journal*, 1972, *51*(9), 662-666.

Glaser, R., Schwarz, P. A., & Flanagan, J. C. Contribution of interview and situational performance to the selection of supervisory personnel. *Journal of Applied Psychology*, 1958, *42*, 69.

Goldberg, L. R. Reliability of Peace Corps selection boards: A study of interjudge agreement before and after board discussions. *Journal of Applied Psychology*, 1966, *50*, 440-448.

Gordon, L. V. Clinical, psychometric and work sample approaches in the prediction of success in Peace Corps training. *Journal of Applied Psychology*, 1967, *51*, 111-119.

Gordon, S. R. The impact of fair employment laws on training. *Training and Development Journal*, November 1978, *32*(11), 29-43.

Grant, D. L., & Bray, D. W. Contributions of the interview to assessment of managerial potential. *Journal of Applied Psychology*, 1969, *53*, 24-34.

Grant, D. L., Katkovsky W., & Bray, D. W. Contributions of projective techniques to assessment of management potential. *Journal of Applied Psychology*, 1967, *51*, 226-232.

Graves, J. P. *Successful and effective management: Contrary effects on subordinate career development*. San Bernardino, CA: Department of Business Administration, California State College, 1978.

Greenhalgh, L. On the evaluation of assessment centers. *Cornell Journal of Social Relations*, 1973, *11*(2), 177-189.

Greenwood, J. M., & McNamara, W. J. Interrater reliability in situational tests. *Journal of Applied Psychology*, 1967, *51*, 101-106.

Hall, H. L., & MacKinnon, D. W. Personality inventory correlates of creativity among architects. *Journal of Applied Psychology*, 1968, *53*, 322-326.

Hardesty, D. L., & Jones, W. S. Characteristics of judged high potential management personnel: The operations of an industrial assessment center. *Personnel Psychology*, *21*(1), 85-98.

Hare, A. P. *Handbook of Small Group Research* (2nd ed.). New York: The Free Press, 1976.

Harlan, A. *Basic Guide for Assessment Interviewing*. Milton D. Hakel, 1971.

Harris, J. G. Prediction of success on a distant Pacific island: Peace Corps style. Journal of Consulting and Clinical Psychology, 1972, 38, 181-190.

Haynes, M. E. Streamlining an assessment center. Personnel Journal, 55(2), 80-83.

Hersey, P. W. NASSP's assessment center. NASSP Bulletin, September 1977, 61(410), 74-88.

Hersey, P.W. NASSP's assessment center--from concept to practice. The Industrial Psychologist, APA Division 14 Publication, 1978.

Hinrichs, J. R. Comparison of real life assessments of management potential with situational exercises, paper and pencil ability tests, and personality inventories. Journal of Applied Psychology, 1969, 53, 425-532.

Hinrichs, J. R. An eight year follow-up of a management assessment center. Journal of Applied Psychology, 1978, 63(5), 596-601.

Hinrichs, J. R., & Haanpera, S. Reliability of measurement in situational exercises: An assessment of the assessment center method. Personnel Psychology, 1976, 29(1), 31-40.

Holmen, M. G. Assessment program of OCS applicants. HUMRRO Technical Report, 1956, 26.

Holt, R. R., & Luborsky, L. Personality patterns of psychiatrists. New York: Basic Books, 1958.

How RPI helps locate talent. Business Week, September 8, 1978, p. 129.

Huck, J. R. Assessment centers: A review of the external and internal validities. Personnel Psychology, 1973, 26(2), 191-212.

Huck, J. R. Determinants of assessment center ratings for white and black females and the relationship of these dimensions to subsequent performance effectiveness. Unpublished doctoral dissertation, Wayne State University, 1974.

Huck, J. R. The Research Base. In J. L. Moses & W. C. Byham (Eds.), Applying the assessment center method. New York: Pergamon Press, 1977.

Huck, J. R., & Bray, D. W. Management assessment center evaluations and subsequent job performance of white and black females. Personnel Psychology, 1976, 29(1), 13-30.

In-Basket Manual. Pittsburgh, PA: Development Dimensions, Inc., 1972.

Jaffee, C. L. A tridimensional approach to management selection. Personnel Journal, 1967, 46, 453.

Jaffee, C. L. Effective management selection: The analysis of behavior by simulation techniques. Reading, MA: Addison-Wesley, 1971.

Jaffee, C. L. Assessment centers. Workshop conducted at the University of Michigan, April 1978.

Jaffee, C. L., Bender, J., & Calvert, D. L. The assessment center technique: A validation study. Management of Personnel Quarterly, 1970, 9(3), 9-14.

Jaffee, C. L., & Frank, F. D. Interviews conducted at assessment centers: A guide for training managers. Dubuque, IA: Kendall/Hunt Publishing Co., 1976.

Jeswald, T. A new approach to identifying administrative talent. The Industrial Psychologist, APA Division 14 Publication, 1978.

Journal of Assessment Center Technology. Winter Park, FL: Assessment Designs, Inc., 1978, 1(1)-2(3).

Kelly, E. L., & Fiske, D. W. Prediction of success in the VA training program in clinical psychology. American Psychologist, 1950, 5, 395-406.

Klimoski, R. J., & Strickland, W. J. Assessment centers: Valid or merely prescient? Personnel Psychology, 1977, 30, 353.

Klinefelter, J., & Capaul, J. Assessment Centers. Sacramento, CA: Selection Consulting Center, 1977.

Klinefelter, J., & Shipe, T. Assessment center approach: A comprehensive summary. Springfield: State of Illinois Department of Personnel, Division of Research and Tests Development, 1972.

Korman, A. K. The prediction of managerial performance: A review. Personnel Psychology, 1968, 21, 295-322.

Kraut, A. I. A hard look at management assessment centers and their future. Personnel Journal, 1972, 51, 317-326.

Kraut, A. I. New frontiers for assessment centers. Personnel, 1976, 53, 30-38.

Kraut, A. I., & Scott, G. J. Validity of an operational management assessment program. Journal of Applied Psychology, 1972, 56, 124-129.

Latham, G. P., Wexley, K. N., & Pursell, E. D. Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 1975, 60, 550-555.

Leadership Decision Styles Survey. Greensboro, NC: Center for Creative Leadership, 1978.

Leary, T. Interpersonal Diagnosis of Personality. New York: The Ronald Press Co., 1957.

Lopez, F. M., Jr. Evaluating executive decision making: The in-basket technique. AMA Research Study 75. New York: AMACOM, 1966.

Lovenheim, B. A Test to Uncover Managerial Skills. The New York Times, January 21, 1979, pp. 1F; 4F.

Lupton, D. E. Assessing the assessment center: A theory and approach. Personnel, 50, 15-22.

MacKinnon, D. W. The use of clinical methods in social psychology. Journal of Social Issues, 1946, 2, 47-54.

MacKinnon, D. W. How assessment centers got started in the United States: The OSS assessment program. Monograph I. Pittsburgh, PA: DDI, 1974.

MacKinnon, D. W. An overview of assessment centers. Technical Report No. 1. Greensboro, NC: Center for Creative Leadership, May 1975.

Marquardt, L. D. Follow-up evaluation of the Second Look Approach to the selection of management trainees. Sears, Roebuck and Company, National Personnel Department, 1976.

Mather, B. E. Assessment of men: A validation study. Unpublished masters thesis, Massachusetts Institute of Technology, 1964.

McCall, M. W., Jr., & Lombardo, N. N. Looking Glass, Inc.: An organizational simulation. Greensboro, NC: Center for Creative Leadership, October 1978.

McConnell, J. H. The assessment center in the smaller company. Personnel, 1969, 46(2), 40-46.

McConnell, J. J., & Parker, T. C. An assessment center program for multi-organizational use. *Training and Development Journal*, March 1972, pp. 6-14.

McCrimmon, M. D. Strategies to increase the cost effectiveness of the assessment center method. *Journal of Assessment Center Technology*, 1978, *1*(2), 13-17.

Meyer, H. H. *Assessment centers at General Electric*. Paper presented at the meeting of Development Dimensions, San Francisco, California, 1972.

Michigan Bell Telephone. *Personnel assessment program: Evaluation study*. Plant Department, 1962.

Mitchel, J. O. Assessment center validity: A longitudinal study. *Journal of Applied Psychology*, 1975, *60*, 573-579.

Morse, J. J., & Wagner, F. R. Measuring the process of managerial effectiveness. *Academy of Management Journal*, 1978, *21*, 23-35.

Moses, J. L. Assessment center performance and management progress. In W. C. Byham (Chair), *Validity of assessment centers*. Symposium presented at the annual convention of the American Psychological Association, Washington, DC, 1971.

Moses, J. L. Assessment center performance and management progress. *Studies in Personnel Psychology*, 1972, *4*, 7-12.

Moses, J. L. The development of an assessment center for the early identification of supervisory potential. *Personnel Psychology*, 1973, *26*(4), 569-580.

Moses, J. L. The assessment center method. In J. L. Moses & W. C. Byham (Eds.), *Applying the assessment center method*. New York: Pergamon Press, 1977.

Moses, J. L. Developing an assessment center program for school administrators. *The Industrial Psychologist*, APA Division 14 Publication, 1978.

Moses, J. L., & Boehm, V. R. Relationship of assessment center performance to management progress of women. *Journal of Applied Psychology*, 1975, *60*, 527-529.

Moses, J. L., & Byham, W. C. (Eds.), *Applying the assessment center method*. New York: Pergamon Press, 1977.

Murray, H. A. *Explorations in personality*. New York & London: Oxford University Press, 1938.

Murray, H. A., & MacKinnon, D. W. Assessment of OSS personnel. *Journal of Consulting Psychology*, 1946, *10*, 76-80.

Myers, D. An assessor's view. *The Industrial Psychologist*, APA Division 14 Publication, 1978.

Nirtaut, D. J. Assessment centers: An examination of participant reaction and adverse effects. *Journal of Assessment Center Technology*, 1978, *1*(2), 18-23.

Office of Strategic Services, OSS Assessment Staff. *Assessment of men*. New York: Rinehart, 1948.

Olmstead, J. A., Cleary, R. K., Lackey, L. L., & Salter, J. A. *Development of leadership assessment simulations*. Arlington, VA: U.S. Army Research Institute for the Behavioral Sciences, Human Resource Research Organization, 1973.

Osgood, C. E., Suci, G. J., & Tannebaum, P. H. *The Measurement of Meaning*. Urbana, IL: University of Illinois Press, 1957.

Parker, T. C. *Summary: A statistical study of the AMA assessment center program*. New York: AMACOM, 1978.

Pomerleau, R. Identification systems: The key to effective manpower planning. *Personnel Journal*, *52*(6), 434-441; 474.

Rice, B. Measuring executive muscle. *Psychology Today*, December 1978, p. 95.

Richards, S. A., & Jaffee, C. L. Blacks supervising whites: A study of interracial difficulties in working together in a simulated organization. *Journal of Applied Psychology*, 1972, *56*, 234-240.

Sackett, P. R., & Hakel, M. D. Temporal stability and individual differences in using assessment information to form overall ratings. *Organization Behavior and Human Performance*, 1979, *23*, in press.

Schwab, D. P., Heneman, H. G., & DeCottis, D. E. Behaviorally anchored rating scales: A review of the literature. *Personnel Psychology*, 1975, *28*, 549-562.

The Seven-Day Leadership Development Program: Information Guide. Greensboro, NC: Center for Creative Leadership.

Shelton, E. *The Adult Performance Level Competency-Based High School Diploma Program*. Austin: The University of Texas at Austin, 1979.

Silzer, R. F. Competence assessment: Use of performance rating scales and assessment centers. Paper presented at the Annual Convention of the American Psycological Association, Toronto, Ontario, 1978.

Silzer, R. F. Development and validation of a center for the assessment of pharmacy competency. Ann Arbor: University of Michigan. College of Pharmacy, 1978.

Simoneit, M. Deutches soldatentum, 1914 und 1939. Berlin: Junker und Buennhaupt, 1940.

Simoneit, M. Grundriss der charakterologischen diagnostic. Leipzig: Teubner, 1944.

Slevin, D. P. The assessment center: Breakthrough in management appraisal and development. Personnel Journal, 51(4), 255-261.

Slivinski, L. W., & Bourgeois, R. P. The inter-rater reliability of the Consolidated Fund In-Basket. Studies in Personnel Psychology, 1974, 6(1), 47-52.

Slivinski, L. W., Grant, K. W., Bourgeois, R. P., & Pederson, L. D. Development and application of a first level management assessment centre. Public Service Commission of Canada, October 1977.

Slivinski, L. W., Grant, K. W., Bourgeois, R. P., Pederson, L. D., & McCloskey, J. L. Longitudinal follow-up of a first-level management assessment center. Journal of Assessment Center Technology, 1978, 1(2), 6-12.

Slivinski, L. W., McCloskey, J. L., & Bourgeois, R. P. Comparison of different methods of assessment. Paper presented to the International Congress on the Assessment Center Method, June 1979.

Smith, L. The EEOC's Bold Foray into Job Evaluation. Fortune, September 11, 1978, pp. 58-61.

State of Wisconsin, Department of Administration: Selection and Career Development Aspects of the Wisconsin Career Executive Program. Unpublished paper, June 1975.

Steiner, R. New use for assessment centers--training evaluation. Personnel Journal, 54(4), 236-237; 248.

Strickland, W. J. The relationship between program evaluation research and selection system validation--application to the assessment center method. Unpublished doctoral dissertation, Ohio State University, 1979.

Taft, R. Multiple methods of personality assessment. Psychology Bulletin, 1959, 56, 333.

Thomson, H. A. Comparison of predictor and criterion judgments of managerial performance using the multitrait-multimethod approach. Journal of Applied Psychology, 1970, 54, 496-502.

Thoresen, J. D., & Jaffee, C. L. A unique assessment center application with some unexpected by-products. Human Resource Management, Spring 1973, 3-7.

Thornton, G. C., III. Varieties of validity of assessment centers in other companies. Proceedings of the 79th Annual Convention, APA, 1971.

Thornton, G. C., III, & McCambridge, J. A. The application of assessment centers to the evaluation and development of graduate students. Unpublished manuscript, Fort Collins: Colorado State University, 1973.

Thornton, G. C., III, & McCambridge, J. A. The use of assessment centers for graduate students. Professional Psychology, November 1974, pp. 366-368.

Vernon, P. E. The validation of Civil Service Selection Board Procedures. Occupational Psychology, 1950(A), 24, 75-95.

Vernon, P. E., & Parry, J. B. Personnel Selection in the British Forces. London: University of London Press, 1949.

Wernimont, P. F., & Campbell, J. P. Signs, samples, and criteria. Journal of Applied Psychology, 1968, 52, 372-376.

Weybright, M. Reactions of an assessee. The Industrial Psychologist, APA Division 14 Publication, 1978.

Wiggins, J. S. Personality and prediction: Principles of personality assessment. Reading, MA: Addison-Wesley Publishing Company, 1973.

Wilson, J. E., & Tatge, W. A. Assessment centers--further assessment needed. Personnel Journal, 1973, 52(3), 172-177.

Wollowick, H. B., & McNamara, W. J. Relationship of components of an assessment center to management success. Journal of Applied Psychology, 1969, 53, 348-352.

Worbois, G. M. Validation of externally developed assessment procedures for identification of supervisory potential. Personnel Psychology, 1975, 28, 71-91.